ระบบคัดกรองกลุ่มเป้าหมายเพื่อการช่วยเหลืออย่างตรงจุดด้วยเทคนิคต้นไม้ตัดสินใจ

A Targeted Assistance Screening System Using Decision Tree Technique

เจษฎา โพนแก้ว[1*], สหัสา พลนิล[2], ธีระพงศ์ สงผัด[3] และ ศุภชัย ทองสุข[4]

Jetsada Ponkaew[1*], Sahutsa Ponnil[2], Teerapong Songputh[3] and Supachai Thongsuk[4]

สาขาวิชาวิทยาการคอมพิวเตอร์[1*] สาขาวิชาเทคโนโลยีคอมพิวเตอร์และดิจิทัล[3] สาขาวิชาวิศวกรรมซอฟต์แวร์[4]

คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยราชภัฏศรีสะเกษ

สาขาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏศรีสะเกษ[2]

Computer Science[1*], Computer and Digital Technology[3], Software Engineering[4],

Faculty of Liberal Arts and Science, Sisaket Rajabhat University

Business Computer, Faculty of Business Administration and Accountancy, Sisaket Rajabhat University[2]

E-Mail : j.ponkaew@sskru.ac.th[1*], sahutsa.p@sskru.ac.th[2], teerapong.song@gmail.com[3], supachai.t@sskru.ac.th[4]

(Received : July 7, 2025; Revised : October 2, 2025; Accepted : December 9, 2025)

บทคัดย่อ

ความยากจนยังคงเป็นปัญหาสำคัญของประเทศกำลังพัฒนา โดยเฉพาะพื้นที่ภาคตะวันออกเฉียงเหนือของประเทศไทยที่มีสัดส่วนครัวเรือนยากจนในระดับสูง ระบบคัดกรองแบบดั้งเดิม เช่น TPMAP ยังขาดความแม่นยำและมักเน้นเพียงข้อมูลเศรษฐกิจ โดยละเลยมิติทางสังคมและวัฒนธรรม ส่งผลให้การจัดสรรทรัพยากรไม่ตรงกลุ่มเป้าหมาย ดังนั้นการวิจัยนี้จึงมีวัตถุประสงค์ 1) พัฒนาโมเดล Decision Tree สำหรับการคัดกรองความยากจนและสกัดกฎ IF-THEN จากข้อมูลครัวเรือนแบบหลายมิติ 2) ประเมินประสิทธิภาพโมเดลเมื่อเปรียบเทียบกับเกณฑ์คัดกรองเดิม (TPMAP) พร้อมวิเคราะห์กฎเชิงตรรกะที่สกัดได้จากโมเดล กลุ่มตัวอย่างคือ ครัวเรือนในจังหวัดศรีสะเกษ จำนวน 1,766 ครัวเรือน รวบรวมจากการสำรวจภาคสนาม และฐานข้อมูลภาครัฐ เครื่องมือวิจัย ได้แก่ แบบสอบถามครัวเรือน ระบบฐานข้อมูล Sisaket Equity System (SES) และโมเดลต้นไม้ตัดสินใจ สถิติที่ใช้ในการวิเคราะห์ ได้แก่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการประเมินด้วย 5-Fold Cross-Validation ร่วมกับตัวชี้วัด Accuracy, Precision, Recall, F1-Score, AUC และ Confusion Matrix

ผลการวิจัยพบว่า 1) โมเดลต้นไม้ตัดสินใจที่พัฒนาขึ้นมีค่า Accuracy เฉลี่ย 0.82 และ AUC เฉลี่ย 0.79 สูงกว่าเกณฑ์ TPMAP ในทุกตัวชี้วัดอย่างมีนัยสำคัญ และ 2) โมเดลสามารถระบุกลุ่มเป้าหมายได้อย่างแม่นยำ ตีความได้ง่าย โดยกฎ IF–THEN ที่สกัดได้แสดงว่าทุนสังคมและทุนมนุษย์เป็นปัจจัยสำคัญที่สุด รองลงมาคือทุนกายภาพ ทุนการเงิน และทุนธรรมชาติ

คำสำคัญ : ความยากจน, การคัดกรองกลุ่มเป้าหมาย, ต้นไม้ตัดสินใจ, ทุนชุมชน, ศรีสะเกษ

ABSTRACT

Poverty remains a critical challenge in developing countries, particularly in northeastern Thailand, where the proportion of poor households is relatively high. Traditional screening systems such as TPMAP still lack accuracy and mainly rely on economic indicators, overlooking social and cultural dimensions, which leads to misallocation of resources and failure to reach the intended target groups. Therefore, this study had two main objectives 1) to develop a decision tree model for multidimensional poverty screening and extract IF–THEN rules from household data, and 2) to evaluate the model's performance in comparison with the existing TPMAP criteria and analyze the logical rules derived from the model. The sample consisted of 1,766 households in Sisaket Province, with data obtained from field surveys and official government databases. The research

instruments included a household questionnaire, the Sisaket Equity System (SES) database, and the developed decision tree model. Descriptive statistics (percentage, mean, and standard deviation) were employed, while model performance was evaluated using 5-fold cross-validation with Accuracy, Precision, Recall, F1-score, AUC, and the confusion matrix as evaluation metrics.

The findings revealed that 1) the developed decision tree model achieved an average accuracy of 0.82 and an average AUC of 0.79, statistically outperforming the TPMAP criteria across all evaluation metrics; and 2) the model was able to identify target households more accurately and in an interpretable manner. The extracted IF–THEN rules indicated that social capital and human capital were the most influential factors in classifying poverty status, followed by physical, financial, and natural capital.

Keywords : Poverty, Target Screening, Decision Tree, Community Capital, Sisaket

บทนำ

ความยากจนเป็นปัญหาเชิงโครงสร้างของประเทศกำลังพัฒนา โดยเฉพาะในเอเชียตะวันออกเฉียงใต้ที่ต้องเผชิญความเหลื่อมล้ำทางเศรษฐกิจ และทางสังคมอย่างต่อเนื่อง [1], [2] ในประเทศไทย แม้เศรษฐกิจเติบโตอย่างต่อเนื่อง แต่ยังมีครัวเรือนราว 6.2 ล้านคนอยู่ภายใต้เส้นความยากจน โดยเฉพาะภาคตะวันออกเฉียงเหนือมีสัดส่วนสูงที่สุด [3], [4] การแก้ปัญหานี้จำเป็นต้องมีกระบวนการคัดกรองกลุ่มเป้าหมายที่แม่นยำ เพื่อให้สามารถจัดสรรทรัพยากรได้อย่างมีประสิทธิภาพ

ระบบคัดกรองแบบดั้งเดิม เช่น TPMAP และบัตรสวัสดิการแห่งรัฐ แม้เป็นกลไกหลักแต่ยังขาดความครอบคลุมโดยเน้นเพียงข้อมูลเศรษฐกิจ และละเลยมิติทางสังคมและวัฒนธรรมซึ่งมีผลต่อความเปราะบางของครัวเรือน [5], [6] ด้วยเหตุนี้งานวิจัยนี้จึงได้ทำการรวบรวมข้อมูลแบบบรูณาการจากหลายแหล่ง รวมถึงข้อมูลจากระบบศรีสะเกษ สุขเสมอกัน [7] ที่พัฒนาขึ้นสำหรับจังหวัดศรีสะเกษ เพื่อให้ครอบคลุมมิติต่าง ๆ ของครัวเรือนอย่างเป็นระบบ

ในปัจจุบันเทคนิค Machine Learning โดยเฉพาะ Decision Tree ได้รับความสนใจในการคัดกรองครัวเรือนยากจน ด้วยข้อดีที่สามารถตีความได้ง่าย และสร้างกฎเชิงเหตุผลได้ชัดเจน [8], [9] เช่น งานวิจัยในเวียดนามรายงานความแม่นยำสูงถึง 85% [10] และใน Costa Rica เทคนิค Random Forest และ Gradient Boosting ก็ให้ผลลัพธ์ดีกว่าเกณฑ์เดิม [11]

งานวิจัยนี้มีจุดเด่นที่แตกต่างจากงานวิจัยก่อนหน้า คือ การบูรณาการข้อมูลจากระบบศรีสะเกษ สุขเสมอกัน [7] แบบหลายมิติที่ครอบคลุมทั้งทุนมนุษย์ ทุนกายภาพ ทุนการเงิน ทุนธรรมชาติ และทุนสังคม [6], [12] เข้ากับเทคนิคต้นไม้ตัดสินใจ [8], [9] ที่สามารถสกัดกฎ IF-THEN ซึ่งมีความง่ายตีความและแปลความหมาย จึงเหมาะนำไปใช้งานได้จริงสำหรับเจ้าหน้าที่ภาคสนาม ซึ่งต่างจากระบบ TPMAP แบบดั้งเดิม ที่เน้นเพียงข้อมูลเศรษฐกิจมิติเดียว [5], [6] และเมื่อเทียบกับเทคนิควิธีป่าไม้แบบสุ่ม (Random Forest) หรือ เครือข่ายใยประสาทเทียม (Neural Networks) [11] ที่แม้จะมีประสิทธิภาพสูง แต่ยากต่อการตีความ และไม่เหมาะกับการใช้งานในพื้นที่จริง

1. วัตถุประสงค์การวิจัย

1.1 เพื่อพัฒนาโมเดล Decision Tree สำหรับการคัดกรองความยากจน และสกัดกฎ IF-THEN จากข้อมูลครัวเรือนแบบหลายมิติ

1.2 เพื่อประเมินประสิทธิภาพโมเดลเมื่อเปรียบเทียบกับเกณฑ์คัดกรองเดิม (TPMAP) พร้อมวิเคราะห์กฎเชิงตรรกะที่สกัดได้จากโมเดล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

ระบบช่วยตัดสินใจ (Decision Support Systems: DSS) เป็นระบบที่ใช้ข้อมูล และโมเดลเพื่อสนับสนุนการแก้ปัญหาที่ซับซ้อนของมนุษย์ โดยมักสกัดองค์ความรู้ในรูปของกฎ IF–THEN จากฐานข้อมูล หรือจากผู้เชี่ยวชาญ การประยุกต์ใช้ระบบเหล่านี้ในงานพัฒนาสังคมช่วยเพิ่มทั้งความโปร่งใสและความน่าเชื่อถือ โดยเฉพาะเมื่อใช้คัดกรองกลุ่มเป้าหมายโครงการสวัสดิการภาครัฐที่ต้องการทั้งความแม่นยำและความยุติธรรม

เทคนิค CART (Classification and Regression Tree) เป็นอัลกอริธึมการเรียนรู้แบบมีโครงสร้าง ซึ่งให้ผลลัพธ์เป็นกฎที่สามารถตีความได้ง่าย และช่วยแยกกลุ่มบุคคลตามสถานะได้อย่างชัดเจน งานของธนาคารโลก [1] และ แองกัส ดีตัน [3] ชี้ให้เห็นว่าการวัดความยากจนในโลกที่เปลี่ยนแปลงต้องใช้เครื่องมือที่ทันสมัยและแม่นยำ งานของ มิลินท์ คชีร์สาการ์ และคณะ [9] ในเวียดนามใช้เทคนิคต้นไม้ตัดสินใจ เพื่อทำนายความยากจนจากข้อมูลดาวเทียมและแบบสำรวจครัวเรือน โดยใช้ฟีเจอร์เพียง 12 ตัวแต่ได้ความแม่นยำสูงถึง 87% ส่วนการศึกษาของ โอลิวิเยร์ ดูปรีเย และคณะ [10] ในเวียดนามยืนยันว่า CART มีประสิทธิภาพสูงในการระบุครัวเรือนที่ต้องการความช่วยเหลือ โดยเฉพาะในพื้นที่ชนบทที่มีข้อจำกัดด้านข้อมูล

งานวิจัยจำนวนมากใช้แบบสอบถามเพื่อวิเคราะห์ความยากจนและคัดกรองกลุ่มเป้าหมาย งานวิจัยในประเทศต่าง ๆ แสดงให้เห็นว่าการใช้ Machine Learning กับข้อมูลแบบสำรวจสามารถทำนายความยากจนได้อย่างมีประสิทธิภาพ โดยงานของ จียุน คิม [11] ในคอสตาริกาใช้ Random Forest และ Gradient Boosting Trees ทำนายความยากจนจากแบบสำรวจ Encuesta Nacional de Hogares พบว่า การศึกษา และการเข้าถึงเทคโนโลยีเป็นปัจจัยสำคัญที่สุด งานเหล่านี้แสดงให้เห็นถึงศักยภาพของการใช้ข้อมูลแบบสอบถามร่วมกับเทคนิค Machine Learning ในการวิเคราะห์ความยากจนได้อย่างมีประสิทธิภาพ

แนวปฏิบัติด้านการพัฒนาชุมชนสมัยใหม่ เน้นการใช้เทคโนโลยีในการคัดกรองแบบมีส่วนร่วม (Technology-Enhanced Community-Based Targeting) เพื่อเลือกกลุ่มเป้าหมายด้านความยากจน งานวิจัยในหลายประเทศแสดงให้เห็นว่าการผสมผสานระหว่าง Machine Learning และความรู้ท้องถิ่นช่วยลดข้อผิดพลาดในการคัดกรองได้อย่างมีนัยสำคัญเมื่อเทียบกับระบบดั้งเดิม โดยเฉพาะการนำเทคนิคดังกล่าวมาใช้ในบริบทการให้ความช่วยเหลือฉุกเฉิน เช่น ในช่วงวิกฤต COVID-19 ตามที่ แอนดรูว์ ซัมเนอร์ และคณะ [2] ประเมินผลกระทบต่อความยากจนโลก ที่หลายประเทศสามารถแจกจ่ายเงินช่วยเหลือผู้ได้รับผลกระทบได้อย่างรวดเร็วและแม่นยำครอบคลุมประชากรจำนวนมาก สอดคล้องกับแนวคิดการใช้ข้อมูลดาวเทียมและโทรศัพท์มือถือตามที่ นีล ฌอง และคณะ [8] ได้ศึกษาไว้

ในประเทศไทย งานด้านการใช้เทคโนโลยีสารสนเทศเพื่อคัดกรองความยากจนเริ่มได้รับความสนใจมากขึ้น สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ [4] และ สถาบันวิจัยเพื่อการพัฒนาประเทศไทย [6] รายงานว่าระบบคัดกรองดั้งเดิมยังมีข้อจำกัด และยังไม่สามารถสะท้อนปัญหาความยากจนได้ครบถ้วนเพียงพอ ข้อค้นพบนี้สอดคล้องกับ อันเดรส กัสตาเญดา อากิลาร์ และคณะ [5] ที่ชี้ว่าการใช้ข้อมูลรายได้เพียงด้านเดียวไม่เพียงพอในการระบุว่าใครคือกลุ่มเป้าหมายที่แท้จริง

เพื่อตอบโจทย์ดังกล่าว งานวิจัยร่วมสมัยจึงหันมาพัฒนาโมเดลเชิงหลายมิติ (Multi-Dimensional Data) ตัวอย่างเช่น มาซารุล โฮก ภูยาน และคณะ [12] ใช้กรอบแนวคิด Community Capital Framework ในบังกลาเทศเพื่อตรวจสอบความยากจนในหลากหลายมิติ และพบว่าทุนทางสังคมและทุนมนุษย์มีอิทธิพลไม่แพ้ทุนทางการเงิน ในประเทศไทย สหัสา พลนิล และคณะ [13] ศึกษาภูมิปัญญาท้องถิ่นด้านงานหัตถกรรมในจังหวัดศรีสะเกษ เพื่อนำองค์ความรู้เชิงวัฒนธรรมมาใช้เสริมการพัฒนา ขณะที่ วิทยา สุขศรี และคณะ [14] ประยุกต์ใช้เทคนิค Data Mining เพื่อวิเคราะห์ข้อมูลผู้ถือบัตรสวัสดิการแห่งรัฐในจังหวัดขอนแก่น และสามารถระบุผู้ที่ควรได้รับการช่วยเหลือด้วยความแม่นยำถึง 82% โดยตัวแปรสำคัญได้แก่ รายได้ครัวเรือน ระดับการศึกษา และสภาพที่อยู่อาศัย

นอกจากนี้ ชัยยา วงศ์ใหญ่ และคณะ [15] ได้พัฒนาระบบสนับสนุนการตัดสินใจโดยใช้ Fuzzy Logic ร่วมกับ Expert System ในการคัดกรองครัวเรือนยากจน ซึ่งช่วยลดระยะเวลาการทำงานของเจ้าหน้าที่ได้ถึง 60%

และเพิ่มความพึงพอใจของผู้ปฏิบัติงาน ส่วน สุภาพร นิลรัตน์ และคณะ [16] ประยุกต์แนวคิดทุนชุมชน 7 ประเภท ในการประเมินความยากจนในลุ่มแม่น้ำโขง โดยใช้โมเดล Random Forest และ XGBoost ซึ่งให้ผลแม่นยำสูงกว่าโมเดลที่ใช้เฉพาะข้อมูลด้านรายได้ถึง 23% งานวิจัยเหล่านี้ยืนยันถึงความสำคัญของการพิจารณาปัจจัยหลายด้าน รวมถึงการนำข้อมูลจาก ระบบศรีสะเกษ สุขเสมอกัน [7] มาประยุกต์ใช้เพื่อการวิเคราะห์เชิงพื้นที่

สำหรับเทคนิค Machine Learning ที่นำมาใช้ในการคัดกรองความยากจนนั้นมีความหลากหลาย Random Forest และ Gradient Boosting (เช่น XGBoost และ LightGBM) เป็นเทคนิคที่ให้ผลลัพธ์แม่นยำสูง ดังเช่นงานของ จียุน คิม [11] และ สุภาพร นิลรัตน์ และคณะ [16] ส่วน Neural Networks ก็มีประสิทธิภาพในการจับความสัมพันธ์ที่ซับซ้อน แต่ข้อจำกัดสำคัญคือความยากในการตีความผลลัพธ์ ตรงกันข้ามกับ Decision Tree ที่แม้ประสิทธิภาพอาจต่ำกว่าเล็กน้อย แต่สามารถสกัดกฎเชิงตรรกะ IF-THEN ที่เข้าใจง่ายและอธิบายได้อย่างโปร่งใส เหมาะสมกับการใช้งานจริงในพื้นที่ ซึ่งเจ้าหน้าที่ต้องการทั้งความแม่นยำและความน่าเชื่อถือในการตัดสินใจจัดสรรสวัสดิการ

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 แบบสอบถามครัวเรือน ครอบคลุม 5 มิติหลัก ได้แก่ ทุนมนุษย์ ทุนกายภาพ ทุนการเงิน ทุนธรรมชาติ และทุนสังคม จัดทำด้วยโครงสร้างที่ชัดเจน และผ่านการตรวจสอบความตรงเชิงเนื้อหา (Content Validity) โดยผู้เชี่ยวชาญจำนวน 3 ท่าน ประกอบด้วย ผู้เชี่ยวชาญด้านการพัฒนาชุมชน ด้านเศรษฐศาสตร์การพัฒนา และด้านสถิติประยุกต์ เพื่อให้มั่นใจว่าเครื่องมือสามารถวัดตัวแปรที่ต้องการได้อย่างครบถ้วนและเหมาะสม

1.2 แบบฟอร์มและไฟล์อิเล็กทรอนิกส์ สำหรับจัดเก็บ ตรวจสอบ และสำรองข้อมูลในรูปแบบดิจิทัลอย่างเป็นระบบ

1.3 เครื่องมือวิเคราะห์ข้อมูล ประกอบด้วยภาษา Python และไลบรารีที่เกี่ยวข้อง ได้แก่ Pandas, Scikit-Learn และ Graphviz พร้อมทั้งประเมินประสิทธิภาพของโมเดลด้วยวิธี K-Fold Cross-Validation, Confusion Matrix และ ROC Curve

2. กลุ่มตัวอย่าง

การวิจัยนี้ดำเนินการในจังหวัดศรีสะเกษ โดยใช้กลุ่มตัวอย่าง 1,766 ครัวเรือนจากทั้ง 22 อำเภอ ที่ได้จากการสอบทาน ข้อมูลครัวเรือนยากจนในฐานข้อมูล TPMAP ปี 2567 ทั้งหมด 28,892 ครัวเรือน ผ่านกระบวนการมีส่วนร่วมร่วมกับ ศจพ.อ. (ศูนย์อำนวยการขจัดความยากจน ระดับอำเภอ) ทุกอำเภอ ข้อมูลถูกรวบรวมจากหลายแหล่ง ได้แก่ TPMAP, ThaiQM, Family Folder และ Sisaket Equity System (SES) เพื่อให้ ครอบคลุมมิติข้อมูลที่หลากหลาย กลุ่มตัวอย่างประกอบด้วยครัวเรือนยากจน 1,539 ราย (87.1%) และครัวเรือนไม่ยากจน 227 ราย (12.9%) การเก็บข้อมูล ได้รับการชี้แจงและความยินยอมจากชุมชน พร้อมการคุ้มครองข้อมูลส่วนบุคคล ตามหลักจริยธรรมการวิจัย

3. ขั้นตอนการดำเนินการวิจัย

การดำเนินการวิจัยประกอบด้วยกระบวนการที่เป็นระบบ เริ่มจากการออกแบบตัวแปรและเครื่องมือวิจัย การรวบรวมและเตรียมข้อมูล ไปจนถึงการสร้างและประเมินประสิทธิภาพของโมเดลต้นไม้ตัดสินใจ กระบวนการทั้งหมดอยู่ภายใต้กรอบแนวคิดเชิงวิชาการ และสอดคล้องกับบริบทของพื้นที่ศึกษาเพื่อให้ได้ผลลัพธ์ที่แม่นยำและนำไปใช้ได้จริง

3.1 การออกแบบตัวแปรและการวัด

เพื่อให้การวิเคราะห์สะท้อนสภาพจริงและสามารถตีความเชิงนโยบายได้อย่างถูกต้อง งานวิจัยนี้ออกแบบตัวแปรตามกรอบแนวคิดทุนชุมชน ครอบคลุม 5 มิติหลัก ได้แก่ ทุนมนุษย์ ทุนกายภาพ ทุนการเงิน ทุนธรรมชาติ และทุนสังคม โดยแต่ละมิติประกอบด้วยฟีเจอร์ย่อยที่สัมพันธ์กับความยากจน เช่น มิติทุนมนุษย์พิจารณาระดับการศึกษา รายได้ สุขภาพ และสวัสดิการ ขณะที่มิติทุนกายภาพพิจารณาความมั่นคงและสภาพของที่อยู่อาศัย มิติทุนการเงินครอบคลุมรายได้รวมต่อปีและการมีหนี้สิน มิติทุนธรรมชาติพิจารณาความมั่นคงของสิทธิในที่ดินและที่อยู่อาศัยในพื้นที่เสี่ยงภัย ส่วนมิติทุนสังคมพิจารณาการเข้าร่วมกลุ่มในชุมชนและการได้รับการสนับสนุนจากเครือข่ายชุมชน

ตัวแปรในงานวิจัยนี้ถูกจำแนกตามประเภทข้อมูลเพื่อให้เหมาะสมต่อการวิเคราะห์ ได้แก่ ตัวแปรเชิงลำดับ (Ordinal) เช่น ระดับการศึกษาและสภาพบ้าน ตัวแปรเชิงหมวดหมู่ (Nominal) เช่น สถานะสุขภาพและความเป็นเจ้าของบ้าน ตัวแปรเชิงปริมาณ (Numerical) เช่น รายได้เฉลี่ยต่อเดือนและรายได้รวมต่อปี และตัวแปรทวิภาค (Binary) สำหรับตัวแปรที่มีเพียงสองค่า เช่น การมีหนี้สิน ความมั่นคงในที่ดิน การอยู่อาศัยในพื้นที่เสี่ยงภัย การเข้าร่วมกลุ่ม และการสนับสนุนจากชุมชน

3.2 การเตรียมและประมวลผลข้อมูล

หลังจากนิยามตัวแปรตามกรอบแนวคิดแล้ว ข้อมูลจะถูกตรวจสอบและจัดการค่าขาดหาย (Missing Values) ด้วยค่าฐานนิยม (Mode) สำหรับตัวแปรเชิงหมวดหมู่ และค่ามัธยฐาน (Median) สำหรับตัวแปรเชิงปริมาณ เพื่อลดผลกระทบจากค่าผิดปกติ (Outliers) ตัวแปรเชิงลำดับเข้ารหัสด้วย Ordinal Encoding ส่วนตัวแปรทวิภาคและเชิงหมวดหมู่แบบสองค่าถูกเข้ารหัสด้วย Binary หรือ Label Encoding

นอกจากนี้ เนื่องจากข้อมูลมีความไม่สมดุลระหว่างกลุ่ม (Class Imbalance) โดยครัวเรือนยากจนมี 1,539 ราย (87.1%) และครัวเรือนไม่ยากจนมี 227 ราย (12.9%) ซึ่งอาจส่งผลต่อประสิทธิภาพของโมเดลในการทำนายกลุ่มน้อย งานวิจัยนี้จึงพิจารณาการจัดการปัญหานี้โดยการตั้งค่า Class_Weight='Balanced' ในอัลกอริทึม Decision Tree ซึ่งจะปรับน้ำหนักของแต่ละคลาสโดยอัตโนมัติ ให้สัดส่วนผกผันกับจำนวนตัวอย่าง ทำให้โมเดลให้ความสำคัญกับทั้งสองกลุ่ม อย่างเท่าเทียมกัน และเพื่อประเมินประสิทธิภาพของโมเดลอย่างครอบคลุม จึงใช้ตัวชี้วัดที่เหมาะสมกับข้อมูลไม่สมดุล ได้แก่ Precision, Recall, F1-Score และ AUC แทนการใช้ Accuracy เพียงอย่างเดียว ซึ่งช่วยให้ ประเมินความสามารถของโมเดลในการระบุทั้งกลุ่มยากจนและไม่ยากจน ได้อย่างเหมาะสม

สำหรับข้อมูลเชิงปริมาณบางตัว ได้ปรับมาตราส่วนด้วย min–max normalization ให้อยู่ในช่วง 0–1 เพื่อให้ตัวแปรมีน้ำหนักเท่าเทียมกัน ข้อมูลทั้งหมดถูกรวบรวมเป็นตารางเชิงโครงสร้าง ครอบคลุมฟีเจอร์สำคัญ 11 ตัว เพื่อฝึกและประเมินโมเดลด้วยวิธี K-Fold Cross-Validation (k=5) โดยแบ่งข้อมูลเป็น 5 ส่วน สลับกันเป็นชุดฝึกและชุดทดสอบจนครบทุก Fold จากนั้นนำผลลัพธ์มาหาค่าเฉลี่ยเพื่อประเมินด้วยตัวชี้วัด ได้แก่ Accuracy, Precision, Recall, F1-Score และ Confusion Matrix เพื่อสะท้อนทั้งความแม่นยำและข้อผิดพลาดของโมเดลอย่างรอบด้าน

3.3 วิธีการสร้างโมเดลต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคการจำแนกและทำนายที่ได้รับความนิยมในงานวิเคราะห์ข้อมูล เนื่องจากโครงสร้างเรียบง่าย ชัดเจน และตีความได้ง่าย โดยเฉพาะในรูปแบบต้นไม้ทวิภาค (Binary Tree) ซึ่งทุกโหนดจะทำหน้าที่แบ่งข้อมูลออกเป็นสองกลุ่มตามเกณฑ์การตัดสินใจที่ดีที่สุด โดยอาศัยค่าเกณฑ์ (Threshold) ของฟีเจอร์ที่เลือก เพื่อให้ได้กลุ่มย่อยที่มีความบริสุทธิ์มากขึ้น วิธีการนี้ช่วยให้เงื่อนไขการตัดสินใจสามารถแสดงในรูปแบบ IF–THEN ที่เข้าใจได้ง่ายและสะท้อนลักษณะของข้อมูลอย่างมีประสิทธิภาพ

ในการสร้างโมเดล แต่ละโหนดจะเลือกฟีเจอร์และค่าเกณฑ์ที่ลดความไม่บริสุทธิ์ของข้อมูลได้มากที่สุด โดยทั่วไปจะใช้ค่า Gini Index (Gini) เป็นตัวชี้วัดความไม่บริสุทธิ์ของโหนด t ซึ่งคำนวณได้จากสมการ:

$$Gini(t) = 1 - \sum_{i=1}^{C}(p(i|t))^2 \quad (1)$$

โดยที่ C คือจำนวนคลาสทั้งหมด และ p(i|t) คือสัดส่วนของตัวอย่างในคลาส i ภายในโหนด t ค่า Gini จะอยู่ในช่วงระหว่าง 0 ถึง 0.5 สำหรับปัญหาการจำแนกแบบสองคลาส โดยค่า 0 หมายถึงโหนดนั้นบริสุทธิ์ (ตัวอย่างทั้งหมดอยู่ในคลาสเดียวกัน) และค่า 0.5 หมายถึงโหนดมีการกระจายคลาสอย่างเท่าเทียมกัน (ซึ่งเป็นกรณีแย่ที่สุด) ในแต่ละการแบ่ง โมเดลจะประเมินว่าการแบ่งข้อมูลไปยังโหนดลูกซ้าย (t_L) และโหนดลูกขวา (t_R) สามารถลดความไม่บริสุทธิ์รวมของกลุ่มย่อยจากโหนดเดิมได้มากน้อยเพียงใด ซึ่งวัดได้จากค่า Reduction in Impurity (RI)

$$RI = i(t) - (\frac{N_L}{N}i(t_L) + \frac{N_R}{N}i(t_R)) \quad (2)$$

โดยที่ i(t) คือค่า impurity ของโหนดเดิม (เช่น Gini) ส่วน i(tL) และ i(tR) คือค่า impurity ของโหนดลูกซ้ายและขวาตามลำดับ และ N_L, N_R คือจำนวนตัวอย่างในโหนดลูกซ้ายและขวาเมื่อเทียบกับจำนวนตัวอย่างทั้งหมด N ค่าของ RI จะอยู่ในช่วงระหว่าง 0 ถึงค่า Gini ของ root_node โดยค่า RI ที่สูง แสดงว่าการแบ่งครั้งนั้นช่วยให้ข้อมูลแยกออกเป็นกลุ่มได้ชัดเจนขึ้นมาก จึงเหมาะสมที่จะเลือกการแบ่งนั้นเป็นเงื่อนไขต่อไป

กระบวนการนี้จะทำซ้ำไปจนกว่าจะถึงเงื่อนไขหยุด เช่น ค่า RI ต่ำกว่าเกณฑ์ที่กำหนด ความลึกของต้นไม้ถึงค่าที่กำหนด (max_depth) หรือจำนวนตัวอย่างในโหนดต่ำกว่าเกณฑ์ขั้นต่ำ (min_samples_split) เมื่อถึงจุดหยุด โหนดนั้นจะกลายเป็น leaf_node และเส้นทางจาก root_node ถึง leaf_node จะนิยามเป็นกฎ IF–THEN สำหรับการตัดสินใจ ตัวอย่างข้อมูลสมมุติจำนวน 4 ครัวเรือนประกอบด้วย ข้อมูลรายได้ การมีไฟฟ้า และสถานะความยากจน (*is_poor*) ดังแสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลครัวเรือนขนาดเล็ก จำนวน 4 ตัวอย่าง

ครัวเรือน	รายได้	ไฟฟ้า	is_poor
1	2000	0	1
2	3000	1	0
3	2500	0	1
4	4000	1	0

ในขั้นแรกจะคำนวณค่า Gini ของโหนดเริ่มต้น (root:is_poor) เพื่อประเมินความไม่บริสุทธิ์ของข้อมูลทั้งหมด โดยครัวเรือนที่จนมี 2 ราย และไม่จนมี 2 ราย ทำให้ได้ค่า

$$Gini(root) = 1 - (0.5)^2 - (0.5)^2 = 0.5$$

ต่อมาทดลองแบ่งข้อมูลโดยใช้เงื่อนไข รายได้ ≤ 2500 ซึ่งแบ่งข้อมูลออกเป็นสองกลุ่ม ได้แก่กลุ่มที่รายได้ ≤ 2500 (ครัวเรือน 1 และ 3) และกลุ่มที่รายได้ > 2500 (ครัวเรือน 2 และ 4) โดยพบว่าในแต่ละกลุ่มย่อยมีความบริสุทธิ์สมบูรณ์ (กลุ่มซ้ายเป็นจนทั้งหมด กลุ่มขวาเป็นไม่จนทั้งหมด) ดังนี้

กลุ่มซ้าย:

$$Gini(left) = 1 - (1)^2 = 0$$

กลุ่มขวา:

$$Gini(right) = 1 - (1)^2 = 0$$

ค่าการลดลงของความไม่บริสุทธิ์ (Reduction in Impurity; RI) จากการแบ่งครั้งนี้จึงคำนวณได้ว่า

$$RI = 0.5 - \left(\frac{2}{4}(0) + \frac{2}{4}(0)\right) = 0.5$$

เนื่องจาก $RI > 0$ และทั้งสองโหนดลูกมีค่า $Gini = 0$ ซึ่งแสดงถึงความบริสุทธิ์เต็มที่ โมเดลจึงหยุดการแบ่งที่นี่ ทำให้ได้เงื่อนไขในการคัดกรองเป็นกฎ IF–THEN ดังนี้

IF รายได้ ≤ 2500 THEN is_poor=1 ELSE is_poor=0

ในกรณีที่มีการใช้หลายฟีเจอร์ในการตัดสินใจ ตัวอย่างเช่น หากพิจารณาพร้อมกันทั้งระดับการศึกษา รายได้เฉลี่ยต่อเดือน และหนี้สิน อาจได้เงื่อนไขในลักษณะ

IF highest_education ≤1 AND avg_monthly_income ≤2500 AND debts=1 THEN is_poor=1 ELSE is_poor=0

โดยกระบวนการคำนวณจะทำซ้ำในทุกโหนดของต้นไม้จนกว่าข้อมูลในโหนดจะบริสุทธิ์ (เช่น $Gini = 0$) หรือเมื่อถึงเงื่อนไขหยุดที่กำหนดไว้ เช่นค่า RI ต่ำกว่าเกณฑ์ ความลึกสูงสุด หรือจำนวนตัวอย่างในโหนดต่ำเกินไป

4. สถิติที่ใช้ในการวิจัย

ในการวิเคราะห์ข้อมูลเชิงพรรณนา ใช้สถิติพื้นฐาน ได้แก่ ร้อยละ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน เพื่ออธิบายลักษณะทั่วไปของครัวเรือน สำหรับการประเมินประสิทธิภาพของโมเดลต้นไม้ตัดสินใจ ใช้ตัวชี้วัด ได้แก่ Accuracy, Precision, Recall, F1-Score และ Confusion Matrix โดยประเมินร่วมกับวิธี K-Fold Cross-Validation เพื่อเพิ่มความน่าเชื่อถือของผลลัพธ์

ผลการวิจัย

ผลการวิจัยนี้นำเสนอตามลำดับวัตถุประสงค์ที่กำหนดไว้ โดยแบ่งเป็น 2 ส่วนหลัก ส่วนแรกเป็นผลการพัฒนาโมเดลต้นไม้ตัดสินใจและการสกัดกฎ IF-THEN จากข้อมูลครัวเรือนแบบหลายมิติ ส่วนที่สองเป็นผลการประเมินประสิทธิภาพของโมเดลเมื่อเปรียบเทียบกับเกณฑ์ TPMAP พร้อมการวิเคราะห์กฎเชิงตรรกะที่สกัดได้จากโมเดล

1. ผลการพัฒนาโมเดล Decision Tree และการสกัดกฎ IF-THEN

1.1 ลักษณะทั่วไปของข้อมูล

ข้อมูลที่ใช้ในการวิจัยนี้มาจากการสำรวจครัวเรือนในจังหวัดศรีสะเกษ จำนวน 1,766 ครัวเรือน ครอบคลุมฟีเจอร์สำคัญ 11 ตัวแปรหลักตามกรอบแนวคิดทุนมนุษย์ ทุนกายภาพ ทุนการเงิน ทุนธรรมชาติ และทุน

สังคม กลุ่มตัวอย่างประกอบด้วยครัวเรือนยากจน 1,539 ครัวเรือน (87.1%) และครัวเรือนไม่ยากจน 227 ครัวเรือน (12.9%) สะท้อนลักษณะข้อมูลที่ไม่สอดคล้องกับสถานการณ์ความยากจนในบริบทจริง

1.2 ผลการเตรียมข้อมูล

หลังนิยามตัวแปรตามกรอบทุนชุมชนแล้ว ได้ตรวจสอบคุณภาพและเตรียมข้อมูลอย่างเป็นระบบ เพื่อให้ครบถ้วน สะอาด และสอดคล้องกับสมมติฐานทางสถิติและกระบวนการเรียนรู้ของเครื่อง จากนั้นจัดทำตารางสรุปฟีเจอร์หลักทั้ง 11 ตัวแปร พร้อมคำจำกัดความและประเภทของตัวแปร ดังแสดงในตารางที่ 2 ซึ่งสะท้อนมิติหลักทั้ง 5 ของทุนชุมชน ได้แก่ มนุษย์ กายภาพ การเงิน ธรรมชาติ และสังคม โดยตารางนี้ทำหน้าที่เป็นเครื่องมือเชื่อมโยงระหว่างทฤษฎีและการปฏิบัติทางเทคนิคในกระบวนการเตรียมข้อมูล

ตารางที่ 2 การกำหนดตัวแปรตามมิติทุนและการแปลงข้อมูล

ชื่อฟีเจอร์	คำจำกัดความ	ประเภทตัวแปร
(1) highest_education	ระดับการศึกษาสูงสุดของหัวหน้าครัวเรือน	Ordinal
(2) avg_monthly_income	รายได้เฉลี่ยต่อเดือนของครัวเรือน	Numerical
(3) health_status	สถานะสุขภาพโดยรวมของครัวเรือน	Categorical
(4) house_ownership	ความเป็นเจ้าของบ้าน (เช่น เช่า/เป็นเจ้าของ)	Categorical
(5) house_condition	สภาพความมั่นคงและความแข็งแรงของบ้านเรือน	Ordinal
(6) household_total_income_per_year	รายได้รวมทั้งปีของครัวเรือน	Numerical
(7) debts	การมีหนี้สิน (1=มี, 0=ไม่มี)	Binary
(8) land_security	ความมั่นคงในการครอบครองที่ดิน (1=มั่นคง, 0=ไม่มั่นคง)	Binary
(9) house_in_disaster_area_past_5yrs	บ้านเรือนตั้งอยู่ในพื้นที่เสี่ยงภัยพิบัติใน 5 ปีที่ผ่านมา (1=ใช่, 0=ไม่)	Binary
(10) participation_in_groups	การเข้าร่วมกลุ่มหรือกิจกรรมของชุมชน (1=เข้าร่วม, 0=ไม่เข้าร่วม)	Binary
(11) community_support_mechanism	การได้รับการสนับสนุนจากกลไกในชุมชน (1=ได้รับ, 0=ไม่ได้รับ)	Binary

จากตารางที่ 2 พบว่า ตัวแปรทั้ง 11 ตัวครอบคลุม 5 มิติของทุนชุมชน ได้แก่ ทุนมนุษย์ (1,2,3) ทุนกายภาพ (4,5) ทุนการเงิน (6,7) ทุนธรรมชาติ (8,9) และทุนสังคม (10,11) โดยผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญ

1.3 ผลการฝึกโมเดล Decision Tree

หลังจากเตรียมข้อมูล ถัดมาเป็นการสร้างโมเดลต้นไม้ตัดสินใจ โดยกำหนดความลึกสูงสุด 5 ชั้น (max_depth=5) และจำนวนข้อมูลขั้นต่ำในการแบ่งกลุ่ม 10 ตัวอย่าง (min_samples_split=10) เพื่อป้องกันการจำแบบอย่างมากเกินไป ใช้เกณฑ์ Gini (Gini Impurity) ในการแบ่งข้อมูล และทดสอบความแม่นยำด้วยการแบ่ง 5-Fold Cross-Validation เพื่อเพิ่มความน่าเชื่อถือของผลลัพธ์ ผลที่ได้แสดงในตารางที่

ตารางที่ 3 ผลการประเมินโมเดล Decision Tree ด้วย 5-Fold Cross-Validation

ตัวชี้วัด	ค่าเฉลี่ย (±SD)
Accuracy	0.82 (±0.03)
Precision	0.84 (±0.02)
Recall	0.89 (±0.04)
F1-score	0.86 (±0.03)
AUC	0.79 (±0.05)

จากตารางที่ 3 แสดงประสิทธิภาพของโมเดล Decision Tree ที่ได้จาก 5-fold cross-validation โดยค่าในวงเล็บแสดงส่วนเบี่ยงเบนมาตรฐาน ค่าตัวชี้วัดที่มากกว่า 0.8 ถือว่าอยู่ในเกณฑ์ดีตามแนวทางทั่วไปในงานด้านการจำแนก

1.4 กฎ IF-THEN ที่สกัดได้

ผลการสกัดกฎ IF-THEN จากโมเดล พบ 4 กฎที่สำคัญซึ่งสะท้อนปัจจัยการจำแนกครัวเรือนยากจนและไม่ยากจนได้อย่างมีนัยสำคัญ ดังนี้:

กฎที่ 1:
IF (community_support_mechanism ≤ 1.5 [ไม่ได้รับการสนับสนุนจากชุมชน]) AND (house_condition ≤ 1.5 [สภาพบ้านไม่มั่นคง]) AND (household_total_income_per_year ≤ 0.50 [รายได้รวมต่อปีต่ำ]) AND (highest_education ≤ 3.02 [การศึกษาต่ำกว่าระดับมัธยมศึกษา]) AND (avg_monthly_income ≤ 0.83 [รายได้เฉลี่ยต่อเดือนต่ำ])
THEN ครัวเรือนจัดอยู่ในกลุ่มยากจน

กฎที่ 2:
IF (community_support_mechanism ≤ 1.5 [ไม่ได้รับการสนับสนุนจากชุมชน]) AND (house_condition > 1.5 [สภาพบ้านมั่นคง]) AND (highest_education > 3.22 [การศึกษาระดับมัธยมศึกษาขึ้นไป]) AND (debts > 1.5 [มีหนี้สิน])
THEN ครัวเรือนจัดอยู่ในกลุ่มไม่ยากจน

กฎที่ 3:
IF (community_support_mechanism > 1.5 [ได้รับการสนับสนุนจากชุมชน]) AND (highest_education ≤ 3.22 [การศึกษาต่ำกว่าระดับมัธยมศึกษา]) AND (participation_in_groups ≤ 3.5 [ไม่ค่อยเข้าร่วมกิจกรรมกลุ่ม])
THEN ครัวเรือนจัดอยู่ในกลุ่มยากจน

กฎที่ 4:
IF (community_support_mechanism > 1.5 [ได้รับการสนับสนุนจากชุมชน]) AND (highest_education > 3.00 [การศึกษาระดับมัธยมศึกษาขึ้นไป]) AND (participation_in_groups > 2.50 [เข้าร่วมกิจกรรมกลุ่มเป็นประจำ]) AND (avg_monthly_income > 0.46 [รายได้เฉลี่ยต่อเดือนปานกลาง]) AND (house_ownership ≤ 1.50 [ไม่ได้เป็นเจ้าของบ้าน])
THEN ครัวเรือนจัดอยู่ในกลุ่มยากจน

2. ผลการประเมินประสิทธิภาพโมเดลเมื่อเปรียบเทียบกับเกณฑ์คัดกรองเดิม (TPMAP) พร้อมวิเคราะห์กฎเชิงตรรกะที่สกัดได้จากโมเดล

2.1 การเปรียบเทียบประสิทธิภาพกับเกณฑ์ TPMAP

เมื่อเปรียบเทียบผลลัพธ์กับโมเดลพื้นฐาน ซึ่งใช้เกณฑ์คัดกรองจากระบบ TPMAP แบบดั้งเดิม พบว่าโมเดล Decision Tree ที่พัฒนาขึ้นสามารถให้ผลลัพธ์ที่ดีกว่าในทุกตัวชี้วัด โดยมีค่า Accuracy และ F1-Score สูงกว่าอย่างมีนัยสำคัญ ซึ่งแสดงถึงศักยภาพของวิธีการที่นำเสนอในการคัดกรองครัวเรือนยากจนได้แม่นยำและมีประสิทธิภาพมากกว่าเกณฑ์เดิม ดังแสดงในตารางที่ 4

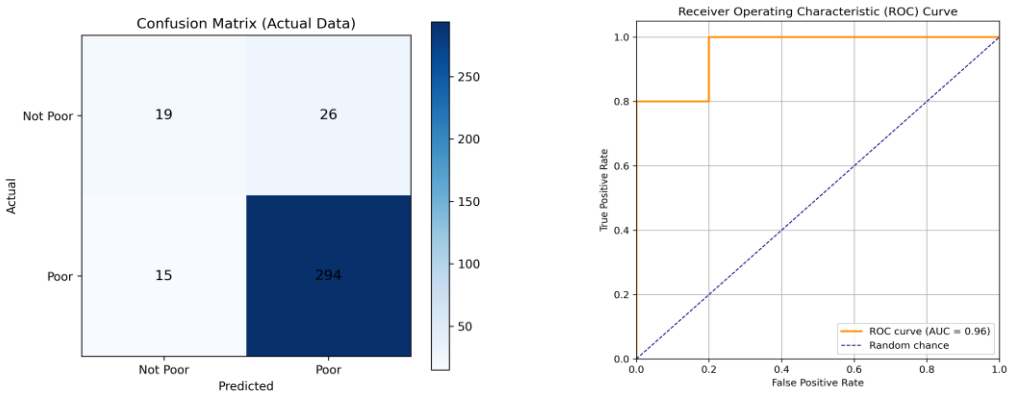
ตารางที่ 4 การเปรียบเทียบผลลัพธ์ระหว่างโมเดล Decision Tree และเกณฑ์ TPMAP

ตัวชี้วัด	Decision Tree	TPMAP (Baseline)
Accuracy	0.82	0.65
Precision	0.84	0.68
Recall	0.89	0.72
F1-score	0.86	0.70
AUC	0.79	0.66

จากตารางที่ 4 แสดงให้เห็นว่า โมเดล Decision Tree มีประสิทธิภาพสูงกว่าเกณฑ์ TPMAP ในทุกตัวชี้วัด โดยมีค่า Accuracy สูงกว่า 17% และ F1-score สูงกว่า 16% อย่างมีนัยสำคัญ

2.2 การประเมินด้วย Confusion Matrix และ ROC Curve

นอกจากการเปรียบเทียบตัวชี้วัดในตารางที่ 4 แล้ว ยังได้มีการประเมินประสิทธิภาพของโมเดลเพิ่มเติม โดยใช้ Confusion Matrix และ (b) ROC Curve ผลดังรูปด้านล่าง



(a) Confusion Matrix (b) ROC Curve

ภาพที่ 1 ผลการประเมินโมเดลด้วย Confusion Matrix และ ROC Curve

จากภาพที่ 1 (a) แสดง Confusion Matrix ของโมเดล ซึ่งสามารถจำแนกครัวเรือนยากจนได้ถูกต้อง (True Positives) จำนวน 294 ครัวเรือน (83.05%) และจำแนกครัวเรือนไม่ยากจนได้ถูกต้อง (True Negatives) อีก 19 ครัวเรือน (5.37%) ขณะเดียวกัน โมเดลมีการทำนายผิดบางส่วน โดยทำนายว่าครัวเรือนไม่ยากจนเป็นยากจน (False Positives) จำนวน 26 ครัวเรือน (7.35%) และทำนายว่าครัวเรือนยากจนเป็นไม่ยากจน (False Negatives) จำนวน 15 ครัวเรือน (4.24%) ผลลัพธ์นี้สะท้อนให้เห็นว่าโมเดลมีประสิทธิภาพในการคัดกรองอยู่ในระดับสูง

ส่วนรูปที่ 1(b) แสดง ROC Curve ของโมเดล โดยเส้นโค้งอยู่เหนือเส้นทแยงมุมอย่างชัดเจน ค่า AUC ที่ได้จากการประเมินเท่ากับ 0.96 บ่งชี้ว่าโมเดลสามารถแยกแยะครัวเรือนยากจนและไม่ยากจนได้อย่างแม่นยำกว่าการทำนายแบบสุ่มอย่างมีนัยสำคัญ

อย่างไรก็ตาม สังเกตได้ว่าค่า AUC ที่แสดงในรูปที่ 1(b) (0.96) สูงกว่าค่าเฉลี่ยที่รายงานในตารางที่ 3 (0.79 ± 0.05) เนื่องจากรูปที่ 1(b) เป็นผลจากการแบ่งข้อมูลเพียงครั้งเดียว (Single Fold) ซึ่งให้ผลลัพธ์ที่ดีที่สุด ขณะที่ค่าเฉลี่ย 0.79 มาจากการประเมินแบบ 5-Fold Cross-Validation ที่สะท้อนภาพรวมของประสิทธิภาพโมเดล

ได้อย่างน่าเชื่อถือกว่า และช่วยลดความเสี่ยงจากการจำข้อมูลมากเกินไป (Over Fitting) ดังนั้น ในการประเมินความแม่นยำของโมเดล ควรอ้างอิงจากค่าเฉลี่ยของ 5-Fold Cross-Validation เป็นหลัก

อภิปรายผลการวิจัย

1. ผลการพัฒนาโมเดล Decision Tree สำหรับการคัดกรองความยากจน และสกัดกฎ IF-THEN จากข้อมูลครัวเรือนแบบหลายมิติ

ผลการวิจัยแสดงว่าเทคนิค Decision Tree สามารถคัดกรองครัวเรือนยากจนได้อย่างมีประสิทธิภาพ โดยโมเดลที่พัฒนาขึ้นให้ค่าตัวชี้วัด (Accuracy, Precision, Recall, F1-Score, AUC) สูงกว่าเกณฑ์คัดกรองเดิม สอดคล้องกับงานของ นีล ฌอง และคณะ [8] และ โอลิวิเยร์ ดูปรีเย และคณะ [10] ที่รายงานว่า Decision Tree มีข้อได้เปรียบด้านการตีความง่ายและแม่นยำเมื่อใช้กับข้อมูลความยากจนในเวียดนามและประเทศอื่น ๆ รวมถึงสอดคล้องกับ มิลินท์ คชีร์สาการ์ และคณะ [9] และ จียุน คิม [11] พบว่า Decision Tree, Random Forest และ Gradient Boosting แยกกลุ่มยากจนได้แม่นยำกว่าเกณฑ์เดิมอย่างมีนัยสำคัญ

กฎ IF-THEN ที่สกัดจากโมเดลชี้ว่าปัจจัยด้านทุนสังคมและทุนมนุษย์มีบทบาทเด่น โดยกฎส่วนใหญ่เริ่มจากเงื่อนไขเกี่ยวกับการสนับสนุนจากชุมชน (Community_Support_Mechanism) และการมีส่วนร่วมในกิจกรรมสังคม (Participation_in_Groups) ผลลัพธ์นี้สอดคล้องกับ มาซารุล โฮก ภูยาน และคณะ [12] และ สหัสา พลนิล และคณะ [13] ที่ชี้ว่าความเข้มแข็งของเครือข่ายสังคมและทุนวัฒนธรรมส่งผลโดยตรงต่อความสามารถในการรับมือกับความยากจน

2. ผลการประเมินประสิทธิภาพโมเดลเมื่อเปรียบเทียบกับเกณฑ์คัดกรองเดิม (TPMAP) พร้อมวิเคราะห์กฎเชิงตรรกะที่สกัดได้จากโมเดล

เมื่อเปรียบเทียบกับระบบ TPMAP พบว่าโมเดล Decision Tree มีประสิทธิภาพเหนือกว่าอย่างชัดเจน โดยค่า Accuracy สูงกว่า 17% และ F1-Score สูงกว่า 16% สะท้อนข้อจำกัดของระบบคัดกรองดั้งเดิมที่มุ่งเน้นเพียงข้อมูลเศรษฐกิจ และละเลยมิติทางสังคมและวัฒนธรรม ซึ่ง อันเดรส กัสตาเญดา อากิลาร์ และคณะ [5] และ สถาบันวิจัยเพื่อการพัฒนาประเทศไทย [6] ต่างชี้ว่าการยกระดับทุนมนุษย์มีบทบาทสำคัญต่อการแก้ปัญหาความยากจนระยะยาว นอกจากนี้ โมเดลยังชี้ให้เห็นว่าระดับการศึกษา (Highest_Education) และรายได้ครัวเรือน (Avg_Monthly_Income) เป็นตัวแปรสำคัญที่สะท้อนศักยภาพการพึ่งพาตนเอง ขณะที่ปัจจัยด้านทุนกายภาพ การเงิน และธรรมชาติ แม้ปรากฏอยู่บ้าง แต่มีบทบาทรองลงมา สอดคล้องกับบริบทพื้นที่ศึกษาที่เครือข่ายทางสังคมและศักยภาพของบุคคลมีอิทธิพลมากกว่า ด้านค่าเฉลี่ยของตัวแปรทั้งห้ามิติยังยืนยันผลนี้ โดยทุนสังคมและทุนมนุษย์ต่ำกว่าเกณฑ์ชัดเจน สอดคล้องกับ แองกัส ดีตัน [3] และ จียุน คิม [11] ที่เน้นความสำคัญของเครือข่ายสนับสนุนและทักษะส่วนบุคคลต่อการยกระดับคุณภาพชีวิต

เมื่อเปรียบเทียบผลการวิจัยนี้กับงานวิจัยในบริบทอื่น พบว่า ค่า Accuracy ที่ 0.82 อยู่ในระดับใกล้เคียงกับงานวิจัยของ โอลิวิเยร์ ดูปรีเย และคณะ [10] ในเวียดนามที่รายงานความแม่นยำ 0.85 แต่สูงกว่างานวิจัยของ นีล ฌอง และคณะ [8] ที่ใช้ข้อมูลดาวเทียมซึ่งให้ค่า 0.78 ความแตกต่างนี้อาจเกิดจากการที่งานวิจัยนี้ใช้ข้อมูลภาคสนามที่รายละเอียดมากกว่า และครอบคลุมหลายมิติของทุนชุมชน รวมถึงการจัดการกับปัญหาข้อมูลไม่สมดุล (Class Imbalance) ด้วยการตั้งค่า Class_Weight='Balanced' ซึ่งช่วยเพิ่มความสามารถในการระบุกลุ่มเป้าหมายได้แม่นยำยิ่งขึ้น นอกจากนี้ การใช้ Decision Tree แทน Random Forest หรือ Gradient Boosting แม้จะให้ประสิทธิภาพต่ำกว่าเล็กน้อย แต่ได้ข้อได้เปรียบด้านความสามารถในการตีความและการสกัดกฎ IF-THEN ที่ชัดเจน ซึ่งมีความสำคัญต่อการนำไปใช้งานจริงในบริบทการคัดกรองสวัสดิการที่ต้องการความโปร่งใสและความเป็นธรรม

ข้อเสนอแนะ

ผลการวิจัยครั้งนี้แสดงให้เห็นว่าเทคนิค Decision Tree ช่วยคัดกรองครัวเรือนยากจนได้แม่นยำและเข้าใจง่าย ควรนำผลลัพธ์ไปใช้พัฒนาระบบคัดกรองให้ทันสมัยและสอดคล้องกับสภาพจริงของพื้นที่ พร้อมทั้งส่งเสริมเครือข่ายชุมชนและจัดอบรมเพิ่มทักษะอาชีพเพื่อช่วยให้ครัวเรือนมีรายได้และความเข้มแข็งมากขึ้น นอกจากนี้ ควรนำกฎที่ได้ไปใช้ในระบบสารสนเทศเพื่อช่วยคัดกรองเป้าหมายได้ง่ายขึ้น ขยายการศึกษาไปยังพื้นที่อื่น ๆ และพัฒนาโมเดลให้แยกความยากจนเป็นด้าน ๆ เช่น สังคม การศึกษา หรือสภาพความเป็นอยู่ เพื่อให้ความช่วยเหลือได้ตรงกับความต้องการมากขึ้น

จากผลการวิจัย มีข้อเสนอแนะเชิงนโยบายดังนี้ 1) หน่วยงาน ที่รับผิดชอบการคัดกรองสวัสดิการควรพิจารณาบูรณาการข้อมูล หลายมิติโดยเฉพาะทุนสังคมและทุนมนุษย์เข้ามาในระบบคัดกรอง ไม่เน้นเพียงรายได้เพียงอย่างเดียว 2) ควรพัฒนาระบบฐานข้อมูล ที่เชื่อมโยงระหว่างหน่วยงาน เช่น เชื่อมโยง SES กับระบบของกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ เพื่อให้การ ตรวจสอบและปรับปรุงข้อมูลเป็นไปอย่างต่อเนื่อง 3) ควรจัดอบรม เจ้าหน้าที่ภาคสนามให้เข้าใจและสามารถใช้กฎ IF-THEN ที่สกัดได้ จากโมเดลในการประเมินเบื้องต้น เพื่อเพิ่มความรวดเร็วและความ สม่ำเสมอในการคัดกรอง 4) ควรมีการติดตามและประเมินผลการช่วยเหลืออย่างเป็นระบบ เพื่อปรับปรุงทั้งโมเดลและแนวทางการ ให้ความช่วยเหลือให้มีประสิทธิภาพมากขึ้น

ข้อจำกัดของงานวิจัย

งานวิจัยนี้แสดงให้เห็นถึงประสิทธิภาพของเทคนิค Decision Tree ในการคัดกรองครัวเรือนยากจนอย่างชัดเจน สำหรับข้อจำกัดที่ควรพิจารณาเมื่อนำไปใช้ในอนาคต ประการแรก ผลการวิจัยนี้มาจากข้อมูล ครัวเรือนในจังหวัดศรีสะเกษ ซึ่งมีบริบททางสังคม เศรษฐกิจ และวัฒนธรรมเฉพาะพื้นที่ การนำโมเดลนี้ไปใช้กับพื้นที่อื่นจำเป็นต้องมีการปรับพารามิเตอร์ หรือฝึกโมเดลใหม่ให้สอดคล้องกับลักษณะเฉพาะของพื้นที่นั้น ๆ ประการที่สอง คือข้อจำกัดของโมเดล Decision Tree ที่ประสิทธิภาพอาจจะด้อยกว่าโมเดลที่ซับซ้อนกว่า เช่น Random Forest หรือ Gradient Boosting ในบางกรณี อย่างไรก็ตาม การเลือกใช้ Decision Tree ในงานวิจัยนี้ เป็นไปเพื่อสมดุลระหว่างประสิทธิภาพ และความสามารถในการนำไปใช้งานจริงของเจ้าหน้าที่ในพื้นที่ ซึ่งต้องการเครื่องมือที่เข้าใจง่าย และสามารถอธิบายเหตุผลการตัดสินใจได้

นอกจากข้อจำกัดด้านบริบทพื้นที่และโมเดลที่กล่าวไปแล้ว งานวิจัยนี้ยังมีข้อจำกัดเพิ่มเติมที่ควรพิจารณา ประการแรก ข้อมูล ที่ใช้ในการวิจัยเป็นข้อมูล ณ จุดเวลาหนึ่ง (Cross-Sectional) ซึ่งไม่ สามารถติดตามการเปลี่ยนแปลงของสถานะความยากจนตามช่วงเวลา ได้การศึกษาแบบติดตามตามยาว (Longitudinal Study) ในอนาคต จะช่วยให้เข้าใจพลวัตของความยากจนและประเมินประสิทธิผลของ การช่วยเหลือได้ดีขึ้น ประการที่สอง ฟีเจอร์ที่ใช้ในโมเดลถูกคัดเลือก จากกรอบแนวคิดทุนชุมชนและข้อจำกัดของข้อมูลที่มีอยู่ อาจมีฟีเจอร์อื่นที่สำคัญแต่ยังไม่ได้รวมอยู่ในโมเดล เช่น ปัจจัยด้านสุขภาพ จิต ความเครียด หรือทุนทางวัฒนธรรม ประการที่สาม ขนาดของกลุ่มตัวอย่างที่ไม่สมดุล (87.1% vs 12.9%) แม้จะได้รับการจัดการ ด้วยเทคนิคทางสถิติแล้ว แต่อาจส่งผลต่อความแม่นยำในการทำนาย กลุ่มไม่ยากจนได้ การเก็บข้อมูลเพิ่มเติมในกลุ่มนี้จะช่วยปรับปรุง ประสิทธิภาพของโมเดลได้

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงด้วยดีจากการสนับสนุนทุนวิจัยโดยหน่วยบริหารและจัดการทุนด้านการพัฒนาระดับพื้นที่ (บพท.) ภายใต้โครงการ “การวิจัยและนวัตกรรมเพื่อยกระดับแพลตฟอร์มขจัดความยากจนแบบเบ็ดเสร็จและแม่นยำ จังหวัดศรีสะเกษ: ระยะที่ 2 ปีที่ 2” ผู้วิจัยขอขอบคุณเจ้าหน้าที่ภาครัฐและหน่วยงานในจังหวัดศรีสะเกษ ตลอดจนครัวเรือนตัวอย่างทุกครัวเรือน ซึ่งเป็นส่วนสำคัญยิ่งต่อความสำเร็จของงานวิจัยนี้

เอกสารอ้างอิง

[1] World Bank, *Poverty and shared prosperity 2022: Correcting course*. Washington, DC: World Bank, 2022. [Online]. Available: https://www.worldbank.org/en/publication/poverty-and-shared-prosperity

[2] A. Sumner, C. Hoy, and E. Ortiz-Juarez, *Estimates of the impact of COVID-19 on global poverty*. United Nations University-WIDER, 2020.

[3] A. Deaton, “Measuring poverty in a growing world (or measuring growth in a poor world),” *Review of Economics and Statistics*, vol. 87, no. 1, pp. 1–19, 2005.

[4] National Economic and Social Development Council, *Thailand poverty statistics report*. Bangkok: NESDC, 2023. (in Thai)

[5] A. Castañeda, D. Doan, D. Newhouse, M. C. Nguyen, H. Uematsu, and J. P. Azevedo, “Who are the poor in the developing world?,” *The World Bank Research Observer*, vol. 33, no. 1, pp. 1–30, 2018.

[6] Thailand Development Research Institute, *Evaluation of the state welfare card program*. Bangkok: TDRI, 2022. (in Thai)

[7] Sisaket Equity System Dashboard, *Sisaket Equity System*. [Online]. Available: https://dashboard.sisaketequity.com/

[8] N. Jean et al., “Combining satellite imagery and machine learning to predict poverty,” *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

[9] M. Kshirsagar, M. Nguyen, and E. Gassner, “Decision trees and poverty mapping: Lessons from Vietnam,” *World Development*, vol. 98, pp. 1–15, 2017.

[10] O. Dupriez, E. Gassner, and M. C. Nguyen, *Using decision trees to improve poverty targeting: Evidence from Vietnam*. World Bank Policy Research Working Paper No. 8396. Washington, DC: World Bank, 2018.

[11] J. Y. Kim, “Using machine learning to predict poverty status in Costa Rican households,” *arXiv preprint* arXiv:2111.13319, 2021. [Online]. Available: https://arxiv.org/abs/2111.13319

[12] M. H. Bhuiyan, M. Rahman, and M. S. R. Sayeef, “The determinants of multidimensional poverty in the urban slums of Dhaka city,” *Regional Sustainability*, vol. 6, no. 1, 2025.

[13] S. Ponnil, P. Thongkham, and K. Thong-on, “Community wisdom and knowledge transfer: A study of banana leaf craft in Sisaket province,” *Journal of Local Development Studies*, vol. 15, no. 2, pp. 45–60, 2021. (in Thai)

[14] W. Sukasri, S. Chantharasatit, and A. Pakdeewanich, “Application of data mining for selecting welfare recipients in Khon Kaen province,” *Journal of Information Technology*, vol. 18, no. 1, pp. 25–38, 2022. (in Thai)

[15] C. Wongyai, K. Nuchanart, and W. Sangtrakoon, “Development of decision support system for beneficiary selection in local development programs using fuzzy logic and expert system,” *Journal of Community Development and Quality of Life*, vol. 11, no. 1, pp. 68–81, 2023. (in Thai)

[16] S. Nilrat, S. Saibua, and W. Sukasri, “Using multidimensional community capital data for poverty assessment in the Mekong River basin,” *Research and Development Journal, Ubon Ratchathani Rajabhat University*, vol. 12, no. 2, pp. 23–38, 2022. (in Thai)