

# การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป

## A Comparison of the Efficiency of Algorithms and Feature Selection Methods for Predicting the Success of Personal Overseas Money Transfer

อนัตต์ชัย ชุตติภาสเจริญ (Anutchai Chutipascharoen)<sup>1</sup>\* ดร.จรัญ แสนราช (Dr.Charun Sanrach)\*\*

### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนาย และคุณลักษณะ ที่มีต่อโอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป โดยทำการศึกษาข้อมูลการโอนเงินของบุคคลทั่วไปจำนวน 51,901 ระเบียบ ทำการเก็บข้อมูลตั้งแต่ปี 2559 - 2560 โดยใช้เทคนิคการจำแนกข้อมูลทั้งหมด 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอิวเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด ซึ่งทำการเปรียบเทียบประสิทธิภาพรูปแบบเทคนิคการทำนาย ระหว่างการใช้คุณลักษณะทั้งหมด การทดสอบประสิทธิภาพด้วยวิธีการ Cross Validation โดยใช้โปรแกรม RapidMiner Studio 8 จากนั้นทำการทดลองเพื่อหาผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้องที่สูงที่สุด ผลการศึกษาพบว่าการใช้เทคนิคต้นไม้ตัดสินใจด้วยการเลือกคุณลักษณะทั้งหมด มีค่าความถูกต้องเท่ากับ 99.90% เทคนิคการค้นหาเพื่อนบ้านใกล้สุดมีค่าความถูกต้องเท่ากับ 99.55% และเทคนิคนาอิวเบย์มีค่าความถูกต้องเท่ากับ 96.71% จากผลการเปรียบเทียบประสิทธิภาพในครั้งนี้ สามารถนำเทคนิคต้นไม้ตัดสินใจ ที่มีค่าความถูกต้องสูงสุดไปใช้ในการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปต่อไป

### ABSTRACT

The study's purpose was to compare the efficiency of algorithms and feature selection methods for predicting the success of personal overseas money transfers. The information of personal overseas money transfers certificate using in their study was 2016-2017 and 51,901 records. This research was used three classification techniques: Decision Tree, Naive Bayes, and K-Nearest Neighbors which compare the efficiency of algorithms the predicting techniques. The efficiency testing can be predicted using Cross Validation method by RapidMiner Studio 8 program. Then, we experimented to find out the result of the efficiency that has maximum accuracy from the study we found that using decision tree technique with attribute. The accuracy is 99.90%, K-Nearest Neighbors 99.55%, and Naive Bayes 96.71%, respectively. Comparison of results we are able to take the technique which has the most accuracy to predict the success of personal overseas money transfers.

**คำสำคัญ:** ต้นไม้ตัดสินใจ นาอิวเบย์ การค้นหาเพื่อนบ้านใกล้สุด

**Keywords:** Decision Tree, Naive Bayes, K-Nearest Neighbors

<sup>1</sup> Correspondent author: S6002042910022@email.kmutnb.ac.th

\*นักศึกษา หลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาคอมพิวเตอร์ศึกษา มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

\*\*ผู้ช่วยศาสตราจารย์ ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ



## บทนำ

การดำเนินธุรกิจในปัจจุบันนี้มีสภาวะการแข่งขันทางธุรกิจที่สูงมากขึ้น และมีการเปลี่ยนแปลงของยุคสมัยที่รวดเร็ว ดังนั้นหลาย ๆ องค์กรจึงต้องมีการหาหลักการหรือวิธีการบริหารจัดการหน่วยงานในองค์กรให้สามารถอยู่รอดในสภาวะการแข่งขันทางธุรกิจได้ และสามารถลดต้นทุนพร้อมทั้งสร้างจุดแข็ง เพื่อเพิ่มศักยภาพการแข่งขันให้เหนือกว่าคู่แข่งอื่น ๆ แต่ยังคงคำนึงถึงผลกำไร และความพึงพอใจสูงสุดของลูกค้าเป็นสำคัญ ซึ่งเทคนิคการพยากรณ์เป็นวิธีการหนึ่งที่ใช้หลักการทางคณิตศาสตร์ และสถิติมาประยุกต์ใช้สำหรับการวางแผนประกอบกับการตัดสินใจในการดำเนินงานทั้งระยะสั้น และระยะยาว นอกจากนี้การพยากรณ์ยังเป็นเครื่องมืออย่างหนึ่งที่จะช่วยให้ได้ข้อมูลสำหรับสร้างความพึงพอใจให้ลูกค้าหรือเพิ่มยอดขายในอนาคต ดังนั้นการพยากรณ์จึงมีบทบาทสำคัญทั้งในการทำงานของภาครัฐ และภาคเอกชน [1]

โดยการคำนวณวิธีการพยากรณ์แต่ละวิธีนั้น ประสิทธิภาพการพิจารณาจากค่าความคลาดเคลื่อนน้อยที่สุดของการพยากรณ์แตกต่างกันไป ซึ่งการเลือกวิธีการพยากรณ์ที่เหมาะสมควรพิจารณาจากหลาย ๆ ปัจจัย ได้แก่ ช่วงการพยากรณ์ที่ต้องการ เวลาที่ใช้ในการพยากรณ์ ลักษณะข้อมูลอนุกรมเวลา และขนาดอนุกรมเวลา สำหรับวิธีการพยากรณ์ที่ได้รับนิยมใช้มีหลายวิธีเช่น เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอิวเบย์ (Naive Bayes) และเทคนิคการค้นหาคู่เพื่อนบ้านใกล้เคียง (k-Nearest Neighbors) เป็นต้น

ในปัจจุบันบริษัทเอกชนที่เกี่ยวข้องกับการให้บริการการเงินข้ามประเทศ ได้ศึกษาการดำเนินธุรกิจในส่วนการ การให้บริการลูกค้า พบว่าปัญหาในเรื่องของการส่งข้อมูล และปัญหาการโอนเงินไม่สำเร็จของลูกค้า อันมีผลมาจากปัญหาทางด้านเทคนิคต่าง ๆ เช่น การรับส่งข้อมูลจากข้อมูลที่เกิดขึ้นในอดีตที่มีความไม่เสถียรภาพเป็นสาเหตุให้ข้อมูลที่ไม่น่าเชื่อถือ ข้อมูลไม่แน่นอน หรือข้อมูลมีความคลุมเครือ ส่งผลทำให้เกิดความคลาดเคลื่อนในการโอนเงินข้ามประเทศ ด้วยเหตุนี้ ผู้วิจัยจึงมีความสนใจที่จะศึกษาปัญหาดังกล่าว และนำมาประยุกต์ใช้สำหรับการพยากรณ์โอกาสความสำเร็จของข้อมูลที่ส่งมาในการทำรายการโอนเงินข้ามประเทศ นอกจากนี้เพื่อศึกษาวิธีการพยากรณ์ที่เหมาะสมสำหรับการนำเสนอโนบายในการวางแผน และการควบคุมการรับส่งข้อมูลบนมาตรฐานเดียวกันอย่างเป็นระบบ ให้มีความถูกต้องแม่นยำมากยิ่งขึ้น ดังนั้นผู้วิจัยจึงมีความสนใจศึกษาการพยากรณ์ที่เหมาะสมกับความต้องการของลูกค้า ใช้วิธีการพยากรณ์ ทั้งหมด 3 รูปแบบ เพื่อลดปัญหาข้อมูลที่มีโอกาสจะส่งโอนเงินข้ามประเทศไม่สำเร็จ และนำเสนอแผนการพยากรณ์ที่เหมาะสม

## วัตถุประสงค์ของงานวิจัย

เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการพยากรณ์โอกาสความสำเร็จในการ โอนเงินข้ามประเทศของบุคคลทั่วไป

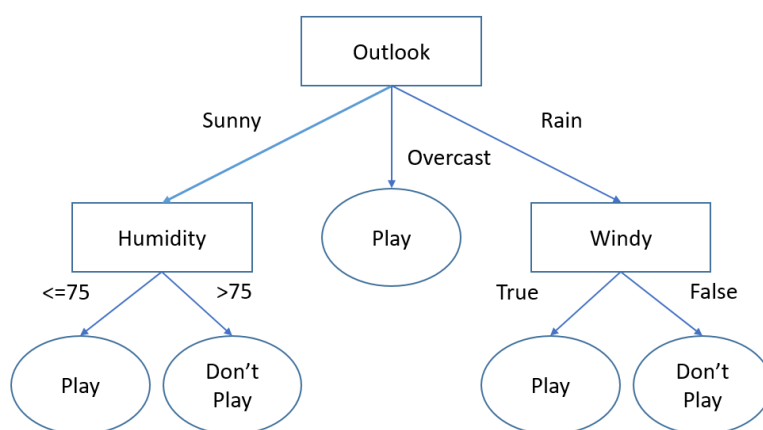
## วรรณกรรมที่เกี่ยวข้อง

การเปรียบเทียบประสิทธิภาพของอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสม เพื่อการพยากรณ์โอกาสความสำเร็จในการ โอนเงินข้ามประเทศของบุคคลทั่วไป ผู้วิจัยได้มีการค้นคว้า และศึกษาข้อมูลต่าง ๆ ที่เกี่ยวข้องเพื่อใช้ประกอบในการจัดทำโครงงาน มีรายละเอียดดังนี้

การจำแนกประเภทข้อมูล (Data Classification) หมายถึง เป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยเป็นการสร้าง โมเดลจำแนกประเภทเพื่อทำนายกลุ่มของข้อมูลใหม่ การสร้างโมเดลจำแนกประเภทข้อมูลเกิดจาก

การหาความสัมพันธ์ของข้อมูล โดยข้อมูลทั้งหมดจะแบ่งเป็น 2 ส่วน คือ ส่วนแรกข้อมูลเรียนรู้ (Training Data) เป็นชุดข้อมูลที่ใช้ในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมาใหม่ เพื่อให้โมเดลที่สร้างได้เรียนรู้ข้อมูล และ ส่วนที่สองของข้อมูลที่ใช้ในการทดสอบโมเดลที่สร้างขึ้นมา (Testing Data) เป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูล เทคนิคพื้นฐานที่นิยมนำมาใช้สำหรับการจำแนกประเภทข้อมูล [2]

ต้นไม้ตัดสินใจ (Decision Tree) หมายถึง กระบวนการทางด้านเหมืองข้อมูลนั้น (Data Mining) ได้นำการตัดสินใจแบบโครงสร้างต้นไม้มาช่วยในการทำงานด้านการตัดสินใจต่าง ๆ ไม่ว่าจะเป็นทางด้านระบบธุรกิจหรือด้านอื่น ๆ โดยปกติจะประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น “If Income = High and Married = No THEN Risk = Poor” โดยลักษณะของการตัดสินใจแบบโครงสร้างต้นไม้มีลักษณะคล้ายกับต้นไม้กลับหัวโดยโหนดแรกสุดจะเป็นรากของต้นไม้ (Root node) แต่ละโหนดแสดงคุณลักษณะ (attribute) แต่ละกิ่งจะแสดงค่าผลในการทดสอบ และ โหนดใบ (Leaf node) [3] การแสดงคลาสที่กำหนดไว้ ดังแผนภาพที่ 1



ภาพที่ 1 แสดงรูปแบบตัวอย่างต้นไม้ตัดสินใจ

อัลกอริทึมนาอิวเบย์ หมายถึง เครื่องจักรเรียนรู้ที่อาศัยหลักการความน่าจะเป็น ตามทฤษฎีของเบย์ (Bayes Theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน เป็นขั้นตอนวิธีในการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอิวเบย์ ใช้การคำนวณหาความน่าจะเป็นในการทำนายผล เป็นเทคนิคในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ จะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์เหมาะกับกรณีของชุดตัวอย่างที่มีจำนวนมาก และคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน โดยกำหนดให้ความน่าจะเป็นของข้อมูลเท่ากับสมการ [4] ดังแผนภาพที่ 2

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

ภาพที่ 2 ตัวอย่างสมการทฤษฎีของเบย์ [5]



การหาเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) หมายถึง วิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่า คลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน “K” ในขั้นตอนวิธีการหาเพื่อนบ้านใกล้ที่สุด ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไข หรือกรณีต่าง ๆ สำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด การนำเทคนิคของขั้นตอนวิธีการหาเพื่อนบ้านใกล้ที่สุดไปใช้นั้น เป็นการหาระยะห่างระหว่างแต่ละตัวแปร (Attribute) ในข้อมูล จากนั้นก็คำนวณค่าออกมา ซึ่งวิธีนี้จะเหมาะสำหรับข้อมูลแบบตัวเลข แต่ตัวแปรที่เป็นค่าแบบไม่ต่อเนื่องนั้นก็สามารถทำได้ เพียงแต่ต้องการการจัดการแบบพิเศษเพิ่มขึ้น อย่างเช่น ถ้าเป็นเรื่องของสี จะใช้อะไรวัดความแตกต่างระหว่างสีน้ำเงินกับสีเขียว ต่อจากนั้นต้องมีวิธีในการรวมค่าระยะห่างของ Attribute ทุกค่าที่วัดมาได้ เมื่อสามารถคำนวณระยะห่างระหว่างเงื่อนไขหรือกรณีต่าง ๆ ได้ จากนั้นก็เลือกชุดของเงื่อนไขที่ใช้จัดคลาส มาเป็นฐานสำหรับการจัดคลาสในเงื่อนไขใหม่ ๆ ได้แล้วจะตัดสินใจได้ว่าขอบเขตของจุดข้างเคียงที่ควรเป็นนั้น ควรมีขนาดใหญ่มากเท่าไร และอาจมีการตัดสินใจได้ด้วยว่าจะนับจำนวนจุดข้างเคียงตัวมัน [6]

หลักการทำงานของ Forward Select หมายถึง การใช้หลักการสร้างชุดของคุณลักษณะโดยการใช้การเรียงลำดับ เริ่มต้นจากเซตว่าง แล้ววนรอบการทำงานโดยเริ่มต้นจากรอบที่ 1 จะทำการหยิบคุณลักษณะทั้งหมดมา 1 คุณลักษณะ เพื่อคำนวณหา Gain ของทุกคุณลักษณะ ซึ่งเป็นค่าที่บอกถึงอำนาจในการจำแนก หากคุณลักษณะใดคำนวณได้ค่า Gain สูง ก็จะเลือกเอาคุณลักษณะนั้นเป็นตัวตั้งสำหรับรอบต่อไป เพื่อทำการจับคู่คุณลักษณะนั้นกับคุณลักษณะอื่น ๆ เพิ่มครั้งละ 1 คุณลักษณะ แล้วก็คำนวณค่า Gain ถ้าคู่ใดได้ค่า Gain สูงก็จะถูกเลือกเป็นตัวตั้งสำหรับรอบถัดไป และวนรอบเพื่อเพิ่มคุณลักษณะขึ้นครั้งละ 1 คุณลักษณะ จนกระทั่งสุดท้ายได้ชุดของคุณลักษณะที่ถูกเลือกออกมา [7]

## วิธีการวิจัย

ผู้วิจัยได้ทำการแบ่งการวิจัยออกเป็น 3 ระยะประกอบด้วย การจัดเตรียมข้อมูล การสร้างตัวแบบทำนายและการวัดประสิทธิภาพของตัวแบบทำนาย มีรายละเอียดดังนี้

ระยะที่ 1 การจัดเตรียมข้อมูล ดำเนินการตามขั้นตอนข้อมูลที่น่ามาใช้ในการวิจัยครั้งนี้ เป็นข้อมูลบุคคลทั่วไปของบริษัทเอกชนจำนวน 51,901 ระเบียน ที่เก็บรวบรวมไว้ตั้งแต่ปี 2559 ถึง 2560 เป็นข้อมูลการโอนเงินไปต่างประเทศทั่วทุกจังหวัด ประกอบด้วยคุณลักษณะทั้งสิ้น 21 คุณลักษณะ ได้แก่ วันที่ เวลา ประเภทการทำธุรกรรม สาขา อาชีพ จังหวัด ประเทศ ประเภทบัตร สัญชาติ จำนวนเงิน ค่าธรรมเนียม จำนวนเงินสุทธิ อัตราแลกเปลี่ยน จำนวนเงินทั้งหมด สถานะการทำงาน รหัสการตรวจสอบ วันที่สร้างใบรายการ จำนวนเงินของผู้รับ ค่าธรรมเนียมของผู้รับ ภาษี จำนวนเงินสุทธิของผู้รับ หลังจากเก็บรวบรวมข้อมูลของแล้ว นำข้อมูลที่ได้มาทำการตรวจสอบความสมบูรณ์ ซึ่งอยู่ในรูปแบบของตาราง Excel และทำการ Clean ข้อมูลเพื่อตัดข้อมูลที่ไม่ว่างออกจนกระทั่งข้อมูลที่มีค่าว่าง (Missing Values) จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ข้อมูลตามอัลกอริทึมที่เลือกใช้

ระยะที่ 2 การสร้างตัวแบบทำนาย เป็นขั้นตอนการออกแบบและสร้างตัวแบบทำนายทั้ง 3 ตัวแบบ ได้แก่ Decision tree, Naive Bayes และ K-Nearest Neighbors ในการวิจัยจะเปรียบเทียบการทำงาน 2 แบบ คือ

1. Simple Classification คือ ไม่มีการเลือกคุณลักษณะของข้อมูลก่อน แต่ใช้อัลกอริทึมในการจำแนกประเภทข้อมูลต่าง ๆ เพื่อทำการประมวลผลเพื่อหาคำตอบอย่างเดียว

2. Hybrid Classification คือ การคัดเลือกคุณลักษณะของข้อมูลก่อนแล้วจึงนำไปประมวลผลต่อในอัลกอริทึมการจำแนกประเภทข้อมูลต่าง ๆ โดยในการวิจัยนี้ใช้เทคนิคการค้นหาคุณลักษณะแบบ Forward Selection

ระยะที่ 3 การวัดประสิทธิภาพของตัวแบบทำนาย ผู้วิจัยได้ทำการทดสอบประสิทธิภาพตัวแบบทำนายด้วยวิธีการ 5 Fold Cross Validation โดยแบ่งข้อมูลเป็น 5 ชุด ไม่ซ้ำกัน ใช้ Train 4 ชุด และ Test 1 ชุด แล้ววนทำซ้ำจนข้อมูลทุกส่วนที่ถูกนำมาทดสอบ สำหรับการพิจารณาที่ดีที่สุด โดยใช้โปรแกรม RapidMiner Studio 8 เป็นเครื่องมือในการวิจัย จากนั้นนำผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้อง (Accuracy) ที่สูงที่สุด 2 ค่า มาทำการเปรียบเทียบ เพื่อตรวจสอบค่าความถูกต้องของชุดข้อมูลทั้งหมด Forward Attribute Decision Tree, Naive Bayes, k-NN ประสิทธิภาพตัวแบบทำนายที่มีค่าความถูกต้องสูงสุด และสุดท้ายได้นำผลของตัวแบบที่ให้ผลดีที่สุดมาใช้สำหรับเป็นต้นแบบในการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปต่อไป

## ผลการวิจัย

การวิจัยนี้ใช้ข้อมูลบุคคลทั่วไปทั่วประเทศไทย จำนวน 51,901 ระเบียน ไปทดสอบกับตัวแบบทำนาย 3 แบบที่ใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอิวเบย์ (Naive Bayes) และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด (k-Nearest Neighbors) โดยวิธีการทดสอบคือ Split test แบ่งออกเป็น 2 ส่วน ส่วนสร้างโมเดลใช้ Training Data 70% ส่วนทดสอบโมเดล Testing Data 30% และผลลัพธ์ของการทำวิจัย ดังตารางที่ 1-5

ตารางที่ 1 Information Gain ของฟีเจอร์ทั้งหมด

| Feature            | Information Gain |
|--------------------|------------------|
| Fee Receive        | 0.000            |
| Tax Receive        | 0.000            |
| Amount Receive     | 0.000            |
| Net Amount Receive | 0.000            |
| Card Type          | 0.001            |
| Register Time      | 0.003            |
| Country            | 0.004            |
| Occupation         | 0.005            |
| Fee                | 0.008            |
| Net Amount         | 0.008            |
| Trans Type         | 0.008            |
| Card Country       | 0.012            |
| Exchange Rate      | 0.013            |
| FC Amount          | 0.013            |
| Check Process      | 0.017            |
| Reg Date           | 0.029            |
| Create Date        | 0.029            |



ตารางที่ 1 Information Gain ของฟีเจอร์ทั้งหมด (ต่อ)

| Feature     | Information Gain |
|-------------|------------------|
| Branch      | 0.032            |
| City        | 0.040            |
| Verify Code | 0.215            |
| Amount      | 0.281            |

จากตารางที่ 1 พบว่าการคัดเลือกฟีเจอร์ทั้งหมด โดยใช้เทคนิคในการคำนวณค่าน้ำหนักของฟีเจอร์ต่าง ๆ โดยฟีเจอร์ที่มีค่าน้ำหนักมากที่สุด คือ Amount มีค่าน้ำหนัก 0.281 ส่วน ฟีเจอร์ที่มีค่าน้ำหนักน้อยที่สุด คือ Fee Receive, Tax Receive, Amount Receive และ Net Amount Receive มีค่าน้ำหนัก 0

ตารางที่ 2 เปรียบเทียบประสิทธิภาพของการเลือกใช้ Forward Selection

| Algorithm              | Accuracy              |                       |
|------------------------|-----------------------|-----------------------|
|                        | Simple Classification | Hybrid Classification |
| 1. Decision tree       | 99.88%                | 99.90%                |
| 2. Naive Bayes         | 22.13%                | 96.71%                |
| 3. K-Nearest Neighbors | 93.29%                | 99.55%                |

จากตารางที่ 2 พบว่าการเลือกคุณลักษณะแบบ Forward Selection มีค่าความถูกต้อง (Accuracy) สูงกว่าการไม่เลือกคุณลักษณะทั้ง 3 เทคนิค โดยเทคนิคที่มีค่าความถูกต้อง (Accuracy) สูงที่สุด คือ เทคนิค Decision Tree มีค่าความถูกต้องเท่ากับ 99.90% เทคนิค K-Nearest Neighbors มีค่าความถูกต้องเท่ากับ 99.55% และเทคนิค Naive Bayes มีค่าความถูกต้องเท่ากับ 96.71%

ตารางที่ 3 ตัววัดประสิทธิภาพของอัลกอริทึม Decision Tree

| Performance Classification (Decision Tree) |        |        |        |        |           |
|--------------------------------------------|--------|--------|--------|--------|-----------|
|                                            | true F | true V | true C | true P | Precision |
| pred. F                                    | 14749  | 3      | 1      | 0      | 99.97%    |
| pred. V                                    | 0      | 786    | 5      | 0      | 99.37%    |
| pred. C                                    | 1      | 0      | 21     | 5      | 77.78%    |
| pred. P                                    | 0      | 0      | 0      | 0      | 0.00%     |
| Recall                                     | 99.99% | 99.62% | 77.78% | 0.00%  |           |

จากตารางที่ 3 พบว่าคลาส F (Finish) มีค่า True Positive (TP) เท่ากับ 14,749 ค่า False Positive (FP) เท่ากับ 4 ค่า Precision 99.97% และ ค่า Recall 99.99% คลาส V (Verify) มีค่า True Positive (TP) เท่ากับ 786 ค่า False Positive (FP) เท่ากับ 5 ค่า Precision 99.37% และ ค่า Recall 99.62% คลาส C (Cancel) มีค่า True Positive (TP) เท่ากับ 21 ค่า False Positive (FP) เท่ากับ 6 ค่า Precision 77.78% และ ค่า Recall 77.78% คลาส P (Pending) มีค่า True Positive (TP) เท่ากับ 0 ค่า False Positive (FP) เท่ากับ 0 ค่า Precision 0% และ ค่า Recall 0%

ตารางที่ 4 ตัววัดประสิทธิภาพของอัลกอริทึม Naive Bayes

| Performance Classification (Naive Bayes) |        |        |        |        |           |
|------------------------------------------|--------|--------|--------|--------|-----------|
|                                          | true F | true V | true C | true P | Precision |
| pred. F                                  | 14529  | 260    | 26     | 5      | 98.04%    |
| pred. V                                  | 221    | 529    | 1      | 0      | 70.44%    |
| pred. C                                  | 0      | 0      | 0      | 0      | 0.00%     |
| pred. P                                  | 0      | 0      | 0      | 0      | 0.00%     |
| Recall                                   | 98.50% | 67.05% | 0.00%  | 0.00%  |           |

จากตารางที่ 4 พบว่าคลาส F (Finish) มีค่า True Positive (TP) เท่ากับ 14,529 ค่า False Positive (FP) เท่ากับ 291 ค่า Precision 98.04% และ ค่า Recall 98.50% คลาส V (Verify) มีค่า True Positive (TP) เท่ากับ 529 ค่า False Positive (FP) เท่ากับ 222 ค่า Precision 70.44% และ ค่า Recall 67.05% คลาส C (Cancel) มีค่า True Positive (TP) เท่ากับ 0 ค่า False Positive (FP) เท่ากับ 0 ค่า Precision 0% และ ค่า Recall 0% คลาส P (Pending) มีค่า True Positive (TP) เท่ากับ 0 ค่า False Positive (FP) เท่ากับ 0 ค่า Precision 0% และ ค่า Recall 0%

ตารางที่ 5 ตัววัดประสิทธิภาพของอัลกอริทึม K-Nearest Neighbors

| Performance Classification (K-Nearest Neighbors) |        |        |        |        |           |
|--------------------------------------------------|--------|--------|--------|--------|-----------|
|                                                  | true F | true V | true C | true P | Precision |
| pred. F                                          | 14726  | 34     | 5      | 0      | 99.74%    |
| pred. V                                          | 21     | 755    | 2      | 1      | 96.92%    |
| pred. C                                          | 3      | 0      | 20     | 4      | 74.07%    |
| pred. P                                          | 0      | 0      | 0      | 0      | 0.00%     |
| Recall                                           | 99.84% | 95.69% | 74.07% | 0.00%  |           |

จากตารางที่ 5 พบว่าคลาส F (Finish) มีค่า True Positive (TP) เท่ากับ 14,726 ค่า False Positive (FP) เท่ากับ 39 ค่า Precision 99.74% และ ค่า Recall 99.84% คลาส V (Verify) มีค่า True Positive (TP) เท่ากับ 755 ค่า False Positive (FP) เท่ากับ 24 ค่า Precision 96.92% และ ค่า Recall 95.69% คลาส C (Cancel) มีค่า True Positive (TP) เท่ากับ



20 ค่า False Positive (FP) เท่ากับ 7 ค่า Precision 74.07% และ ค่า Recall 74.07% คลาส P (Pending) มีค่า True Positive (TP) เท่ากับ 0 ค่า False Positive (FP) เท่ากับ 0 ค่า Precision 0% และ ค่า Recall 0%

จากการเปรียบเทียบประสิทธิภาพของอัลกอริทึมทั้ง 3 แบบพบว่าการใช้เทคนิค Decision Tree ด้วยการเลือกคุณลักษณะแบบ Forward Selection ค่าความถูกต้อง (Accuracy) ในการจำแนกประเภทข้อมูลได้ดีที่สุด โดยให้ค่าความถูกต้อง ที่ 99.90% สามารถสรุปเป็นกฎการตัดสินใจ แสดงดังแผนภาพที่ 3

```

AMOUNT > 167.945
| FEE > 75
| | EXCHANGERATE > 0.004: F {F=2400, V=0, C=0, P=0}
| | EXCHANGERATE ≤ 0.004
| | | CHECKPROCESS = false: C {F=0, V=0, C=3, P=0}
| | | CHECKPROCESS = true: V {F=0, V=33, C=1, P=0}
| FEE ≤ 75: F {F=12349, V=1, C=1, P=0}
  
```

ภาพที่ 3 ตัวอย่างการสร้างกฎจากต้นไม้ตัดสินใจ

จากภาพที่ 3 พบว่าจากผลลัพธ์ของการทำงานในของเงินนั้นจะแบ่งออกได้ 4 สถานะ คือ สถานะ Finish หมายถึง สถานะการทำการโอนเงินสำเร็จ สถานะ Pending หมายถึง สถานะการทำการโอนเงินต้องรอการยืนยันจากสำนักงานใหญ่อีกครั้ง สถานะ Verify หมายถึง สถานะการทำการโอนเงินถูกส่งกลับไปที่สาขาต่าง ๆ ที่ทำการมา เพื่อให้กลับไปแก้ไขข้อมูลให้ถูกต้อง และสถานะ Cancel หมายถึง สถานะการทำการโอนเงินถูกยกเลิกเนื่องจากสาเหตุความผิดปกติของข้อมูล เช่น การติดบัญชีดำหรืออื่น ๆ กฎการตัดสินใจที่ได้ คือ IF Amount > 167.945 AND FEE ≤ 75 OR FEE > 75 AND EXCHANGERATE > 0.004 หมายความว่า ถ้าจำนวนในการส่งเงินมากกว่า 167.945 บาท และค่าธรรมเนียมน้อยกว่าหรือเท่ากับ 75 บาท หรือถ้าค่าธรรมเนียมมากกว่า 75 บาท อัตราแลกเปลี่ยนต้องมากกว่า 0.004 มีโอกาสที่จะส่งเงินสำเร็จสถานะเป็น Finish และถ้าส่งจำนวนเงินน้อยกว่าหรือเท่ากับ 167.945 มีโอกาสที่จะมีสถานะเป็น Verify และ Cancel ถ้าข้อมูล Check Process=True จะเป็นสถานะ Verify และถ้า Check Process = False จะเป็นสถานะ Cancel

## สรุปผลการศึกษา

การเปรียบเทียบประสิทธิภาพของอัลกอริทึม และการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไป โดยในการดำเนินงานเริ่มต้นตั้งแต่การศึกษาข้อมูลของเทคนิคการทำเหมืองข้อมูล วิธีในการ Training Data รูปแบบ Testing Data ซึ่งการวิเคราะห์สังเคราะห์จากข้อมูล และเอกสารกำหนดขอบเขตของ ทำการออกแบบแผนการทดลอง โดยใช้โปรแกรม RapidMiner Studio 8 ในการสร้างโมเดลตามเทคนิคต่าง ๆ ในการทดลอง ซึ่งช่วยในการเขียนโมเดลโดยใช้เทคนิคการ Cleansing Data การ Select Data รวมไปถึงการทำ Data Mining และประมวลผลผลลัพธ์ ทำให้ระยะเวลาในการทดลองมีความรวดเร็ว และมีประสิทธิภาพมากยิ่งขึ้น หลังจากการทำเหมืองข้อมูลเสร็จเรียบร้อยแล้วสามารถนำผลจากการวิจัยในการพยากรณ์โอกาสความสำเร็จในการโอน

เงินข้ามประเทศของบุคคลทั่วไป ได้ทั้ง 3 ตัวแบบ ประกอบไปด้วย Decision tree, Naive Bayes และ K-Nearest Neighbors ทำการเปรียบเทียบเพื่อให้ได้วิธีการที่ได้ค่าความถูกต้องที่สูงที่สุด จากนั้นจึงทำการเปรียบเทียบประสิทธิภาพความถูกต้องในการทำนาย เพื่อหาตัวแบบทำนายที่ให้ค่าความถูกต้องสูงสุดที่เหมาะสมกับการนำไปใช้ในการทำนายโอกาสความสำเร็จในการโอนเงินข้ามประเทศ ผลการวิจัยพบว่าตัวแบบการทำนายที่ใช้เทคนิค Decision tree ร่วมกับเทคนิคการเลือกคุณลักษณะ Forward Selection ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 99.90 เทคนิค K-Nearest Neighbors ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 99.55 และเทคนิค Naive Bayes ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 96.71 ตามลำดับสามารถบอกได้ว่าโอกาสความสำเร็จของข้อมูลที่ส่งมามีโอกาสส่งเงินสำเร็จหรือไม่ และสามารถนำไปเขียนโปรแกรมเพื่อให้ระบบมีการตรวจสอบข้อมูลได้อย่างถูกต้องมากยิ่งขึ้น และลดข้อผิดพลาดของข้อมูลให้ลดน้อยลง ซึ่งจะทำให้การโอนเงินข้ามประเทศมีโอกาสสำเร็จมากขึ้น จากการวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพอัลกอริทึม และการคัดเลือกคุณลักษณะที่เหมาะสม เพื่อการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปในครั้งนี้ ทำให้ทราบว่าหากต้องการทำเหมืองข้อมูลที่มีปริมาณมาก ๆ ต้องมีอุปกรณ์ที่มีความเร็วสูง เพื่อให้สามารถสนับสนุนการทำงานได้อย่างเต็มประสิทธิภาพ และลดการเกิดข้อผิดพลาดต่าง ๆ ในระหว่างการประมวลผลข้อมูล นอกจากเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอิวเบย์ (Naive Bayes) และเทคนิคการค้นหาเพื่อนบ้านใกล้เคียง (k-Nearest Neighbors) ที่ใช้ในการทดลองแล้วควรจะศึกษาทฤษฎีที่เกี่ยวข้อง และหลักการอื่น ๆ ที่ผู้วิจัยไม่นำมาประยุกต์ใช้เพิ่มเติม เพื่อเพิ่มประสิทธิภาพด้านต่าง ๆ ขึ้น

## เอกสารอ้างอิง

1. Ruekkasaem L. Demand Forecasting for Production Planning : A Case Study of Cleanroom Apparel. Parichart J Thaksin Univ. 2015; 28(3): 290–305.
2. Hengphraphrom K, Hengphraphrom S, Makwibunchai S. The Knowledge Discovery of Students' Critical Characteristics towards Learning Achievement of Nakhon Pathom Rajabhat University Students' Program in Computer via Data Mining Technique. J Thai Interdiscip Res. 2014; 9(1): 71–80.
3. Quinlan JR. Induction of decision trees. Mach Learn. 1986; 1(1): 81–106.
4. Vilailuck S, Jaroenpuntaruk V, Wichadakul D. Utilizing Data Mining Techniques to Forecast Student Academic Achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center. Veridian E-J Sci Technol Silpakorn Univ. 2015; 2(2): 1–17.
5. Charoenvorakiat P. Bayes' theorem [Internet]. ARTIFICIAL INTELLIGENCE, DATA MINING, FEATURE, INSPIRATION, QUANTUM COMPUTER. 2016. Available from: <http://www.parinya.net/node/2295>
6. Thongjor N. Machine Learning#2 [Internet]. Babel Coder. 2017. Available from: <https://www.babelcoder.com/blog/posts/k-nearest-neighbors>
7. Chaveesuk R, Juthajaroenwong T. Demand forecasting and farm production planning for frozen further processed broilers: A case study of an integrated operation [Internet]. Vol. 2. Kasetsart University, Bangkok (Thailand); 2010. Available from: [http://kukr.lib.ku.ac.th:80/db/BKN/search\\_detail/exportBibTex/313263](http://kukr.lib.ku.ac.th:80/db/BKN/search_detail/exportBibTex/313263)