

แบบจำลองการทำนายแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายอัตโนมัติ

Models for automatic aircraft type prediction

¹ภูมิพัฒน์ ดวงกลาง ²รัชนี เครือแก้ว

¹ภาควิชาคอมพิวเตอร์ กองวิชาคณิตศาสตร์และคอมพิวเตอร์ กองการศึกษา โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช

²กองการศึกษา โรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช

¹phummipat_d@rtaf.mi.th, ²rajana@rtaf.mi.th

บทคัดย่อ

Received: January 2, 2019

Revised: February 13, 2019

Accepted: March 11, 2019

การป้องกันภัยทางอากาศเป็นหนึ่งในภารกิจที่สำคัญของกองทัพอากาศ โดยขั้นตอนแรกของการปฏิบัติคือการค้นหาวัตถุที่มีความเร็วบนท้องฟ้าที่อาจเป็นภัยคุกคามต่อความมั่นคงของประเทศด้วยการใช้เรดาร์ สิ่งที่เราตรวจพบนั้น ในขั้นต้นถูกเรียกว่าเป้าหมายไม่ทราบฝ่าย หรือก็คือวัตถุบินได้ที่ยังไม่ผ่านกระบวนการพิสูจน์ทราบว่าเป็นอากาศยานแบบใดและเป็นของฝ่ายใด การวิเคราะห์เพื่อระบุแบบอากาศยานและฝ่ายอย่างถูกต้องนี้เป็นปัจจัยสำคัญในการเลือกหนทางปฏิบัติของกองทัพอากาศเพื่อตอบสนองให้เหมาะสมกับสถานการณ์ เช่น หากเป็นอากาศยานรบของฝ่ายตรงข้าม กองทัพอากาศก็จะจัดอากาศยานรบสมรรถนะสูงขึ้นบินสกัดกั้น และแจ้งเตือนหน่วยต่อสู้อากาศยานให้พร้อมต่อสู้ป้องกันฐานบินด้วยยุทธวิธีที่เหมาะสม เป็นต้น

ปัจจุบันการวิเคราะห์และระบุเป้าหมายไม่ทราบฝ่าย กระทำโดยเจ้าหน้าที่ผู้เชี่ยวชาญที่มีประสบการณ์ในการทำงานและการฝึกร่วมกับกองทัพอากาศมิตรประเทศ ความแม่นยำของการวิเคราะห์ขึ้นอยู่กับขีดความสามารถของเจ้าหน้าที่แต่ละคน ซึ่งการผลิตเจ้าหน้าที่ให้มีขีดความสามารถและเชี่ยวชาญในงานด้านนี้ต้องใช้เวลาหลายปี ประกอบกับขีดความสามารถดังกล่าวมีความเกี่ยวข้องกับการให้บริการการบินที่หน่วยงานพลเรือนภายนอกที่ให้ค่าตอบแทนสูงมีความต้องการ เป็นเหตุให้การขอลาออกจากราชการของบุคลากรกลุ่มนี้เพื่อไปประกอบอาชีพกับหน่วยงานเหล่านั้นเกิดขึ้นอย่างต่อเนื่อง

เพื่อแก้ไขปัญหาดังกล่าว ผู้วิจัยได้สร้างแบบจำลองการทำนายแบบอากาศยานอัตโนมัติที่ไม่พึ่งพามนุษย์ โดยใช้ข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านของกรมข่าวทหารอากาศเป็นข้อมูลในการวิจัยเพื่อค้นหาแบบจำลองที่เหมาะสม ด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) ได้แก่ Naïve Bayes, Deep Learning, Decision Tree และ Random Forest ซึ่งผลการวิจัยพบว่า แบบจำลองที่เหมาะสมที่จะนำไปใช้งานกับกระบวนการวิเคราะห์และระบุเป้าหมายไม่ทราบฝ่ายมากที่สุด คือแบบจำลองที่ได้จากเทคนิค Random Forest เนื่องจากมีความถูกต้องในการทำนายผลสูง ผู้วิจัยเชื่อมั่นว่าแบบจำลองดังกล่าวจะสามารถเพิ่มประสิทธิภาพให้กระบวนการตัดสินใจของผู้บังคับบัญชากระทำได้รวดเร็วและเหมาะสมยิ่งขึ้น อีกทั้งผลที่ได้จากงานวิจัยนี้จะสามารถแก้ปัญหาการขาดแคลนบุคลากรผู้เชี่ยวชาญ และลดการพึ่งพาความชำนาญของเจ้าหน้าที่เฉพาะทางลงได้ สอดคล้องกับแนวความคิดการใช้ระบบอาวุธที่ฉลาด (Smart Weapon Systems) ที่นำไปสู่การเลือกกลยุทธ์ที่ฉลาด (Smart Tactics) ของกองทัพอากาศอีกด้วย

คำสำคัญ : เหมืองข้อมูล, Random Forest, Decision Tree, Deep Learning

Abstract

Air defense is one of the major missions of the Air Force. The first step is to find objects with speed which could pose a threat to national security by using radar. Basically, any object first detected by radar is determined to be an unknown target. In other word, it is a flying object which type and owner have not yet been identified. An accurate analysis of aircraft type and owner is an important factor for the decision on how the RTAF responds to the situation. For example, if it's an opponent's combat aircraft, the RTAF will be able to intercept and alert aircraft units to defend the air base with appropriate tactics.

At present, the analysis and identification of unknown targets are being handled by expert staffs with experience in operating and training with allied forces. The accuracy of the analysis depends on the capabilities of each officer. Training the staff to be competent and specialized in this field takes several years. The capability is also relevant to the aviation services that external agencies require, where the compensation is much higher. As a result, the resignation of the personnel of this group to sign contract with such agencies continuously occurs.

To solve this issue, the researcher constructed a model for predicting unknown aircraft using the data from the air traffic analysis system of neighboring countries owned by Directorate of Intelligence by implementing data classification technique i.e. the Bayesian Network, SVM, and Decision Tree. The best result was a model derived from the Random Forest technique, because of its high accuracy in prediction. The researcher believes that this model can improve the decision-making process of the commanders more quickly and appropriately. The results of this research will be able to solve problems of expert personnel shortage and reduce the dependence on the expertise of specialized staffs in accordance with the concept of Smart Weapon Systems, which lead to the selection of the Smart Tactics of the RTAF.

Keywords : Data Mining, Random Forest, Decision Tree, Deep Learning

1. บทนำ

การวิเคราะห์เป้าหมายไม่ทราบฝ่ายคือการวิเคราะห์ข้อมูลที่ตรวจจับได้จากเรดาร์ของระบบเฝ้าตรวจทางอากาศเทียบกับฐานข้อมูลข่าวกรองความเคลื่อนไหวกำลังทางอากาศ ยุทธวิธีการฝึก และแบบอากาศยานที่มีประจำการในกองทัพอากาศประเทศรอบบ้านและเวียดนาม (เท่าที่รัศมีเรดาร์ของกองทัพอากาศ (ทอ.) และบริษัทวิทยุการบินประเทศไทยจำกัดครอบคลุมไปถึง) ซึ่งการวิเคราะห์เป้าหมายไม่ทราบฝ่ายนั้นถือเป็นการกิจที่สำคัญต่อการป้องกันภัยทางอากาศของประเทศ โดยปัจจุบันกรมข่าวทหารอากาศ (ขว.ทอ.) เป็นหน่วยงานรับผิดชอบในการเก็บ

ผลลัพธ์ที่ผ่านการวิเคราะห์หาแบบอากาศยานแล้วลงในระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้าน ซึ่งระบบดังกล่าวฯ ได้พัฒนาขึ้นด้วยวัตถุประสงค์เพื่อปรับโครงสร้างการจัดเก็บผลลัพธ์การวิเคราะห์แบบอากาศยาน ซึ่งเดิมอยู่ในรูปแบบเอกสารให้เป็นระบบฐานข้อมูลที่สามารถสืบค้น และปรับปรุงข้อมูลได้อย่างเป็นระบบและมีประสิทธิภาพ อีกทั้งยังสามารถใช้จัดความสามารถในการประมวลผลของระบบคอมพิวเตอร์เพิ่มความน่าเชื่อถือในการวิเคราะห์และผลิตข้อมูลให้ตรงกับความต้องการใช้งานของผู้ใช้ข้อมูลแต่ละกลุ่มมากยิ่งขึ้น ทำให้สามารถสนับสนุนวงรอบการตัดสินใจ (OODA Loop) ของผู้บังคับบัญชา

ในการใช้กำลังทางอากาศให้มีความรวดเร็วและเหมาะสมมากยิ่งขึ้น

กระบวนการวิเคราะห์ข้อมูลเป้าหมายไม่ทราบฝ่ายที่ตรวจพบจากเรดาร์เฝ้าตรวจของกองทัพอากาศนั้นประกอบด้วยขั้นตอน 4 ขั้นตอน และมีหน่วยเกี่ยวข้อง 2 หน่วย คือ กรมข่าวทหารอากาศ (ขว.ทอ.) ซึ่งเป็นเจ้าของระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้าน และศูนย์ยุทธการทางอากาศ ศูนย์ปฏิบัติการกองทัพอากาศ (ศยอ.สปก.ทอ.) ซึ่งรับผิดชอบการปฏิบัติงานในการวิเคราะห์และระบุแบบอากาศยานจากข้อมูลเป้าหมายไม่ทราบฝ่ายที่ตรวจพบด้วยเรดาร์เบื้องต้นในภาพรวม ผู้วิจัยขออธิบายโดยสังเขป ดังนี้

ขั้นตอนการวิเคราะห์แบบอากาศยานขั้นต้น เป็นการเริ่มต้นกระบวนการทำงานของการวิเคราะห์เป้าหมายไม่ทราบฝ่าย โดยเจ้าหน้าที่วิเคราะห์ข้อมูลที่ปฏิบัติงานในศยอ.สปก.ทอ. จะศึกษาข้อมูลและรูปแบบการบินของอากาศยานที่แสดงบนจอภาพของระบบบัญชาการและควบคุมทางอากาศ (Air Command and Control System: ACCS) ที่เชื่อมโยงเรดาร์เฝ้าตรวจของกองทัพอากาศและบริษัทวิทยุการบินฯ มาวิเคราะห์เพื่อหาแบบอากาศยานที่คาดว่าจะเป็นอย่างใด ซึ่งการวิเคราะห์ในขั้นตอนนี้จะใช้เพียงค่าคุณลักษณะพื้นฐานของเป้าหมายทางอากาศที่ได้รับจากสัญญาณเรดาร์ เช่น ความเร็ว ความสูง ทิศทาง และพิกัด เป็นต้น เมื่อได้ผลลัพธ์แล้วจะบันทึกลงไฟล์ข้อมูลในรูปแบบตาราง (Spreadsheet) แล้วนำส่งให้กับ ขว.ทอ. เพื่อดำเนินการในขั้นตอนต่อไป

ขั้นตอนการตรวจสอบความถูกต้อง เป็นขั้นตอนการตรวจสอบและวิเคราะห์ข้อมูลเพิ่มเติมเพื่อให้ข้อมูลมีความถูกต้องมากยิ่งขึ้น ซึ่งเป็นหน้าที่ของนักวิเคราะห์ข้อมูลในส่วน of ขว.ทอ. เนื่องจากข้อมูลที่ได้รับมาจาก ศยอ.สปก.ทอ. เป็นข้อมูลผ่านการวิเคราะห์ทางเทคนิค โดยใช้เพียงค่าคุณลักษณะพื้นฐานที่ได้รับจากเรดาร์เท่านั้น ซึ่งในทางข่าวกรองถือเป็นข้อมูลจากแหล่งเดียว มีความเป็นไปได้ที่จะคลาดเคลื่อนไปจากข้อเท็จจริงสูง ดังนั้นเจ้าหน้าที่วิเคราะห์ข้อมูลในส่วน of ขว.ทอ. จะต้องตรวจสอบความถูกต้องโดยการวิเคราะห์ข้อมูลเพิ่มเติมร่วมกับข้อมูลด้านการข่าวจากแหล่งอื่นๆ และเมื่อวิเคราะห์

และตรวจสอบข้อมูลเรียบร้อยแล้ว จะส่งต่อให้เจ้าหน้าที่บันทึกข้อมูลลงในระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านต่อไป

ขั้นตอนการนำข้อมูลเข้าระบบฯ หลังจากข้อมูลได้ผ่านการตรวจสอบความถูกต้อง เจ้าหน้าที่จะบันทึกข้อมูลดังกล่าวเข้าสู่ระบบวิเคราะห์ความเคลื่อนไหวฯ ต่อไป

ขั้นตอนการวิเคราะห์ข้อมูลเชิงสถิติ เป็นขั้นตอนการเตรียมข้อมูล และการวิเคราะห์ข้อมูลเชิงสถิติสำหรับรายงานผู้บังคับบัญชาระดับสูง โดยระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศของประเทศรอบบ้านได้เตรียมฟังก์ชันการวิเคราะห์และการแสดงผลในรูปแบบของแผนภูมิต่าง ๆ ที่เหมาะสมกับวัตถุประสงค์และความต้องการใช้งานที่แตกต่างกัน และผลลัพธ์ที่ได้จากการประมวลผลข้อมูลทางสถิตินี้ ยังมีความจำเป็นต้องพึ่งพาเจ้าหน้าที่ข่าวกรองอาวุโสที่มีความเชี่ยวชาญในการแปลผลเพื่อผลิตข่าวกรองยุทธวิธี และข่าวกรองยุทธศาสตร์ทางทหารของประเทศนั้น ๆ

ระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านมีฟังก์ชันการทำงานที่มีประสิทธิภาพ แต่การปฏิบัติงานโดยรวมยังต้องพึ่งพาเจ้าหน้าที่วิเคราะห์ข้อมูลที่มีความชำนาญสูงเป็นหลัก ซึ่งบุคลากรกลุ่มดังกล่าวเป็นบุคลากรที่กองทัพอากาศกำลังประสบปัญหาขาดแคลน และยังมีแนวโน้มที่ผู้ปฏิบัติงานในปัจจุบันจะลาออกจากราชการไปทำงานกับหน่วยงานภายนอกมากขึ้นทุกปี การผลิตเจ้าหน้าที่ทดแทนก็ต้องใช้เวลามาก ไม่ทันต่อความต้องการใช้งานของกองทัพอากาศ

เพื่อแก้ไขปัญหาดังกล่าว การทำเหมืองข้อมูล (Data Mining) ซึ่งเป็นเทคนิคที่ใช้สำหรับการค้นหาค้นหาความรู้ที่ซ่อนอยู่ภายในฐานข้อมูลที่เก็บสะสมไว้ สำหรับใช้วิเคราะห์วางแผน ทำนายอนาคต รวมถึงการตัดสินใจนั้น เมื่อนำมาใช้สร้างแบบจำลองช่วยในการวิเคราะห์และทำนายแบบอากาศยาน จึงนับเป็นทางเลือกที่เหมาะสมกับปัญหาและสถานการณ์

กระบวนการการทำเหมืองข้อมูลแม้จะถูกนำไปประยุกต์ใช้งานอย่างแพร่หลายในหลายสาขา โดยเฉพาอย่างยิ่งภาคธุรกิจ แต่ในกองทัพอากาศนั้น ยังไม่พบว่ามีมีการประยุกต์ใช้กระบวนการทำเหมืองข้อมูลเพื่อวิเคราะห์ข้อมูลหรือใช้เพิ่มขีดความสามารถให้กับระบบงานใด ๆ ของกองทัพอากาศมาก่อน

ดังนั้น งานวิจัยนี้จะเป็นการทดลองและศึกษาการประยุกต์ใช้การทำเหมืองข้อมูลกับระบบงานของกองทัพอากาศ โดยนำข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศของประเทศรอบบ้านมาสร้างแบบจำลองการทำนายแบบอากาศยานด้วยเทคนิคการทำเหมืองข้อมูล นอกจากนี้ยังมีการเปรียบเทียบประสิทธิภาพความแม่นยำในการทำนายผลและเวลาในการประมวลผลของเทคนิคแต่ละแบบ เพื่อพิจารณาหาแบบจำลองที่เหมาะสมต่อการนำไปประยุกต์ใช้งานต่อไป

2. ขอบเขตงานวิจัย

การวิจัยนี้เป็นการทดลองสร้างแบบจำลองทำนายแบบอากาศยาน รวมถึงเปรียบเทียบประสิทธิภาพการทำนายผลของแบบจำลองจากแต่ละเทคนิคการทำเหมืองข้อมูล ในประเด็นของความถูกต้องของการวิเคราะห์ขั้นต้น

3. ทฤษฎีที่เกี่ยวข้อง

การทำเหมืองข้อมูล (Data mining) [1] คือ กระบวนการหรือเทคนิคที่ใช้สำหรับการค้นหาคำตอบที่ซ่อนอยู่ภายใต้ฐานข้อมูลขนาดใหญ่ เพื่อใช้สำหรับการวิเคราะห์ วางแผน ทำนายอนาคต ตลอดจนช่วยในการตัดสินใจ ซึ่งการประยุกต์ใช้การทำเหมืองข้อมูลให้มีประสิทธิภาพ จำเป็นอย่างยิ่งที่จะต้องศึกษาประเภทของข้อมูล รวมถึงเทคนิคการวิเคราะห์แต่ละแบบ เพื่อที่จะสามารถเลือกเทคนิคการวิเคราะห์ให้สอดคล้องและเหมาะสมกับข้อมูลที่จะนำมาวิเคราะห์ โดยข้อมูลหลักๆ จะมีอยู่ด้วยกัน 2 รูปแบบ ได้แก่ ข้อมูลแบบมีโครงสร้าง (Structured data) ซึ่งคือข้อมูลที่มีโครงสร้างที่ชัดเจน โดยทั่วไปจะอยู่ในรูปแบบของตาราง ข้อมูลรูปแบบนี้เป็นข้อมูลที่มีความเหมาะสมสำหรับการวิเคราะห์และประมวลผล และข้อมูลแบบที่สองคือข้อมูลแบบไม่มีโครงสร้าง (Unstructured data) เป็นข้อมูลที่ไม่มีความแน่นอน ได้แก่ ข้อความ หรือรูปภาพ เป็นต้น ซึ่งข้อมูลแบบที่สองนี้จะต้องแปลงรูปแบบของข้อมูลให้อยู่ในรูปแบบที่มีโครงสร้างก่อน เพื่อให้เกิดความสะดวกและเหมาะสมต่อการวิเคราะห์

สำหรับเทคนิคการวิเคราะห์ข้อมูลจะสามารถจำแนกออกเป็น 2 กลุ่มหลัก คือ

กลุ่มที่ 1 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) เป็นเทคนิคที่เน้นการพิจารณาเพื่อค้นหาความสัมพันธ์กันของข้อมูลเป็นหลัก ซึ่งเทคนิคการวิเคราะห์ข้อมูลประเภทนี้จะแบ่งย่อยได้เป็น 2 กลุ่ม ได้แก่ เทคนิคการค้นหากฎความสัมพันธ์ (Association rule) และการแบ่งกลุ่มข้อมูล (Clustering)

กลุ่มที่ 2 การเรียนรู้แบบมีผู้สอน (Supervised learning) จะเป็นเทคนิคที่อาศัยการเรียนรู้จากข้อมูลที่มีอยู่ในอดีตเพื่อนำมาสร้างแบบจำลองสำหรับการทำนายหรือคาดการณ์อนาคต โดยเทคนิคนี้จะใช้การจำแนกประเภทข้อมูล (Classification) เพื่อสร้างรูปแบบหรือแบบจำลองในการใช้หาคำตอบที่ต้องการจะทำนาย ดังนี้

(1) Naïve Bayes เป็นเทคนิคที่ใช้ความน่าจะเป็นของข้อมูลมาช่วยในการสร้างแบบจำลองในการทำนายผลลัพธ์ โดยอ้างอิงทฤษฎีของ Bayes [2] มีข้อดีคือง่ายและสะดวกต่อการนำไปพัฒนาใช้งาน แต่ข้อเสียคือ เหมาะกับข้อมูลทดสอบที่มีคุณลักษณะอิสระจากกัน [3]

(2) Deep Learning เป็นการนำเทคโนโลยีโครงข่ายสมอง (Neural Network) เข้ามาช่วยในการสร้างแบบจำลองการทำนายผลลัพธ์ โดยการสร้างระดับชั้นการเรียนรู้ ซึ่งความรู้แต่ละระดับชั้นจะมีการแยกแยะองค์ประกอบพื้นฐานความรู้ รวมถึงแต่ละระดับชั้นจะมีความสัมพันธ์กันที่ซับซ้อนเหมือนโครงข่ายสมอง [4]

(3) Decision Tree คือเทคนิคการสร้างแบบจำลองที่ใช้ช่วยในการตัดสินใจ โดยแบบจำลองจะอยู่ในรูปแบบโครงสร้างต้นไม้ (Tree) โดยมีใบ (leaf) เป็นสิ่งที่อยู่ล่างสุด บ่งบอกถึงชุดข้อมูลคำตอบ (class) ซึ่งเป็นผลลัพธ์สุดท้ายที่ผ่านการทดสอบเงื่อนไขตามคุณลักษณะของแต่ละโหนด (node) ตามเส้นทางของกิ่งต้นไม้ (branch) [5]

(4) Random Forest เป็นการสุ่มเลือกข้อมูลเพื่อนำมาทำ Decision Tree หลายๆ แบบจำลองเพื่อทำนายผล ซึ่งจำนวนผลลัพธ์ที่ซ้ำกันมากที่สุดของแต่ละแบบจำลองจะเป็นคำตอบของการทำนายผล [6]

4. การดำเนินการวิจัย

การสร้างแบบจำลองสำหรับวิเคราะห์และทำนายแบบอากาศยานจากข้อมูลของระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านนั้น จะอ้างอิงกระบวนการมาตรฐาน CRISP-DM [7] ดังนี้

4.1 ทำความเข้าใจปัญหา (Business Understanding)

จากการศึกษาขอบเขตของงานวิจัย และกระบวนการวิเคราะห์เป้าหมายไม่ทราบฝ่าย จะสามารถกำหนดวัตถุประสงค์ของงานวิจัยได้คือ การสร้างแบบจำลองทำนายแบบอากาศยานสำหรับเป็นเครื่องมือช่วยเจ้าหน้าที่วิเคราะห์ข้อมูลในส่วนของ ขว.ทอ. ในการตรวจสอบความถูกต้องของผลลัพธ์ที่ได้จากการวิเคราะห์แบบอากาศยานขั้นต้นของ จนท.ประจำจอเรดาร์ของ ศยอ.สปก.ทอ.

4.2 ทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลคือปัจจัยที่สำคัญที่สุดสำหรับการทำเหมืองข้อมูล ซึ่งในขั้นตอนนี้จะเป็นการรวบรวมข้อมูลที่เกี่ยวข้องสำหรับการวิเคราะห์ โดยควรพิจารณาแหล่งที่มาของข้อมูล ความน่าเชื่อถือ และจำนวนข้อมูลมีปริมาณมากพอต่อการนำไปใช้วิเคราะห์ เมื่อพิจารณาแหล่งที่มาของข้อมูลของการวิจัยขั้นนี้ ถือได้ว่าแหล่งข้อมูลมีความน่าเชื่อถือ เนื่องจากข้อมูลในระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านเป็นข้อมูลของกรมข่าวทหารอากาศที่ใช้ในการผลิตข่าวกรองสรุปนำเรียนผู้บังคับบัญชาเป็นประจำทุกเดือน และข้อมูลดังกล่าวมีจำนวนมากกว่า 37,000 ระเบียบ ซึ่งเพียงพอต่อการนำมาวิเคราะห์ด้วยกระบวนการทำเหมืองข้อมูล โดยมีตัวแปรที่จะนำมาวิเคราะห์รายละเอียดตามตารางที่ 1

ตารางที่ 1 รายการตัวแปรสำหรับสร้างแบบจำลอง

ตัวแปร	ประเภท	ช่วงข้อมูล [Min-Max]
NumberOfPlane	Integer	[1 - 7]
Velocity	Integer	[2 - 33000]
FoundHeight	Integer	[9 - 49000]
FoundTime	Time	[0:00 – 23:59]
FoundLon	Real	[89 - 117.583]
FoundLat	Real	[-10 - 30.2]

TakeOffAirport	Polynomial	[รายชื่อสนามบิน 35 แห่ง]
LandedAirport	Polynomial	[รายชื่อสนามบิน 39 แห่ง]
HideTime	Time	[0:00 – 23:59]
HideHeight	Integer	[3 - 36000]
HideLon	Real	[85 - 119.917]
HideLat	Real	[0.067 - 29.75]
CountryName	Polynomial	[Malaysia, Myanmar, Laos, Cambodia]
PlaneName	Polynomial	[Fighter, Transporter, Training, Helicopter]

4.3 เตรียมข้อมูล (Data Preparation) ขั้นตอนนี้เป็นการพิจารณาคัดเลือกตัวอย่างข้อมูล สำหรับนำมาใช้เป็นชุดข้อมูลสำหรับสร้างแบบจำลองทำนายแบบอากาศยาน ในเบื้องต้นต้องทำการคัดเลือกข้อมูลที่มีความสมบูรณ์ (No Missing value) รวมถึงคัดกรองข้อมูลที่อยู่ภายนอกกลุ่มออกไป (Outlier) โดยใช้เทคนิค Euclidean distance จนได้ผลลัพธ์คือชุดข้อมูลที่เหมาะสมกับการสร้างแบบจำลอง โดยจำนวนชุดข้อมูลที่สมบูรณ์และเหมาะสมต่อการสร้างแบบจำลองมีจำนวน 6,440 ระเบียบ แบ่งเป็น 4 คลาสคำตอบประเภทแบบอากาศยาน (PlaneName) ได้แก่ อากาศยานประเภทขับไล่โจมตี (Fighter) จำนวน 2,991 ระเบียบ , อากาศยานประเภทลำเลียง (Transporter) จำนวน 2000 ระเบียบ, เฮลิคอปเตอร์ (Helicopter) จำนวน 1,000 ระเบียบ และ อากาศยานใช้สำหรับการฝึก (Training) จำนวน 449 ระเบียบ

แต่เนื่องจากจำนวนข้อมูลในแต่ละคลาสมีจำนวนที่แตกต่างกันมากเกิดความไม่สมดุลของข้อมูล (Imbalance) ซึ่งจะส่งผลกระทบต่อประสิทธิภาพการทำนายผลของแบบจำลอง ดังนั้นผู้วิจัยจึงได้ทำการแก้ไขโดยใช้วิธี Under Sampling เพื่อลดจำนวนข้อมูลของแต่ละคลาสคำตอบให้มีจำนวนที่ใกล้เคียงกันมากที่สุด โดยจำนวนข้อมูลในแต่ละคลาสคำตอบหลังจากการทำ Under Sampling และจะถูกใช้เป็นชุดข้อมูลสำหรับการสร้างแบบจำลอง คือ แบบอากาศยานประเภทขับไล่โจมตี (Fighter) จำนวน 500 ระเบียบ, อากาศยานประเภทลำเลียง (Transporter) จำนวน 500 ระเบียบ เฮลิคอปเตอร์ (Helicopter)

จำนวน 500 ระเบียน และ อากาศยานใช้สำหรับการฝึก (Training) จำนวน 449 ระเบียน รวมทั้งสิ้น 1,949 ระเบียน

4.4 สร้างแบบจำลอง (Modeling) ขั้นตอนนี้เป็นการสร้างแบบจำลองด้วยเทคนิคการจำแนกประเภทข้อมูล (Classification) โดยใช้ชุดข้อมูล (Training Dataset) จำนวน 1949 ระเบียนที่ได้จากขั้นตอนที่ผ่านมา ซึ่งการดำเนินการมีรายละเอียด ดังนี้

1) ทำการ Normalization ข้อมูลที่เป็นตัวเลข ได้แก่ NumberOfPlane, Velocity, FoundHeight, FoundLon, FoundLat, HideHeight, HideLon, HideLat เนื่องจากข้อมูลของ Attributes เหล่านี้มีช่วงความกว้างของข้อมูลที่แตกต่างกันมาก ซึ่งอาจจะส่งผลกระทบต่อประสิทธิภาพของแบบจำลอง

2) ประมวลผลสร้างแบบจำลองโดยใช้เครื่องคอมพิวเตอร์แม่ข่ายที่มีคุณลักษณะ คือ หน่วยประมวลผลกลาง (CPU) แบบ 8 แกนหลัก (8 core) และมีหน่วยความจำหลัก (RAM) ชนิด ECC DDR4 จำนวน 32 GB.

3) ใช้โปรแกรม Rapid miner 9 ในการสร้างแบบจำลอง

4) แบ่งกลุ่มข้อมูลเพื่อทดสอบประสิทธิภาพการทำนายผลของแบบจำลองแบบ Cross-Validation Test โดยใช้ค่า $k = 10$

5) ใช้เทคนิค Classification โดยเลือกวิธีการ Naïve Bayes, Deep Learning, Decision Tree และ Random Forest สำหรับสร้างแบบจำลองทำนายแบบอากาศยาน โดยมีการกำหนดค่าของแต่ละวิธีการ ดังนี้

ในส่วนของ Deep Learning การทดลองนี้จะเลือกใช้แบบจำลอง Deep Learning แบบ Multi-Layer Feed-Forward Artificial Neural Network โดยกำหนดค่าตัวแปร Activation เป็น Exponential Rectifier Linear Unit function (ExpRectifier) รวมถึงกำหนด Hidden Layer เป็น 100, 200, 100 ซึ่งหมายถึงแบบจำลองจะมี Hidden Layer จำนวน 3 ชั้น และชั้นที่สองจะมีจำนวน neurons เท่ากับ 200 ส่วนชั้นที่แรกและชั้นที่สามจะมีจำนวน neurons เท่ากับ 100

สำหรับ Decision Tree และ Random Forest ได้กำหนดค่า Critetion = Accuracy และ Maximal Depth=10

นอกจากนี้ยังมีการกำหนดค่า Number of Trees = 100 เพิ่มเติมให้กับวิธีการ Random Forest อีกด้วย

4.5 การประเมินผล (Evaluation) การประเมินผลเป็นการประเมินความสามารถของแบบจำลองในการทำนายแบบอากาศยาน ซึ่งหากผลการประเมินค่าของแบบจำลองไม่เหมาะสมต่อการนำไปใช้ประโยชน์ จำเป็นต้องย้อนกลับไปขั้นตอนก่อนหน้า เพื่อเปลี่ยนแปลงแก้ไขชุดข้อมูลทดสอบให้ได้ผลลัพธ์ตามที่ต้องการ ดังนั้นเพื่อให้ได้แบบจำลองทำนายแบบอากาศยานที่แม่นยำและมีประสิทธิภาพ โดยการพิจารณาค่าในตำแหน่งต่าง ๆ ของ Confusion Matrix รายละเอียดตามรูปที่ 1 ดังนี้

TP (True Positive) คือ ผลการทำนายถูกต้องตามค่าที่คาดหวัง

FP (False Positive) คือ ผลทำนายไม่ถูกต้อง เนื่องจากข้อมูลจริงที่ไม่อยู่ในคลาส แต่ผลทำนายว่าอยู่ในคลาส

TN (True Negative) คือ ผลการทำนายถูกต้อง สำหรับความจริงที่ข้อมูลไม่อยู่ในคลาส และผลทำนายว่าไม่อยู่ในคลาส

FN (False Negative) คือ ผลการทำนายไม่ถูกต้อง ซึ่งความเป็นจริงข้อมูลอยู่ในคลาส แต่ในขณะที่ผลทำนายว่าไม่อยู่ในคลาส

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

รูปที่ 1 รูปแบบของ Confusion Matrix

จากรายละเอียด Confusion Matrix สามารถนำมาคำนวณหาค่ามาตรวัดคุณภาพเพื่อประเมินประสิทธิภาพของแบบจำลอง ดังนี้

1) Precision คือ ความแม่นยำในการทำนายผลของแบบจำลองโดยพิจารณาในแต่ละคลาสคำตอบ โดยใช้สูตรคำนวณ $Precision = TP / (TP + FP)$

2) Recall คือ ค่าความสามารถในการทำนายของแบบจำลองที่ตรงกับความเป็นจริง ซึ่งจะพิจารณาแต่ละคลาสคำตอบ และจะใช้สูตรคำนวณ $Recall = TP / (TP + FN)$

3) Accuracy คือ ค่าความเที่ยงตรง หรือ ค่าความถูกต้องในการทำนายผลของแบบจำลอง โดยพิจารณาข้อมูลจากทุกคลาสคำตอบ ซึ่งคำนวณได้จาก $Accuracy = (TP + TN) / \text{จำนวนข้อมูลทั้งหมด}$

4.6 การนำไปใช้งาน (Deployment) เป็นการประยุกต์ใช้แบบจำลองการทำนายแบบอากาศยานที่มีค่าความถูกต้องสูงและใช้เวลาในการประมวลผลที่สามารถยอมรับได้ร่วมกับกระบวนการวิเคราะห์เป้าหมายไม่ทราบฝ่ายในขั้นตอนการตรวจสอบความถูกต้องของผลลัพธ์

5. ผลการวิจัย

หลังจากได้แบบจำลองที่ผ่านการประมวลผลจากการใช้ชุดข้อมูล Training DataSet ร่วมกับเทคนิค Naïve Bayes, Deep Learning, Decision Tree และ Random Forest แล้ว จึงได้มีการทดสอบประสิทธิภาพความแม่นยำในการทำนายผล และผลการทดลองได้ดังตารางที่ 2 ซึ่งจะเห็นได้ว่ามีเพียง 2 แบบจำลอง คือ Random Forest และ Decision Tree ที่มีค่าผลการทำนายที่สูงเกินร้อยละ 80 ทั้ง 3 มาตรวัดคุณภาพ โดยแบบจำลองของ Random Forest มีค่า Accuracy, Precision และ Recall มากถึงร้อยละ 90.2, 90.47 และ 90.24 ตามลำดับ ส่วน Decision Tree จะมีค่า Accuracy, Precision และ Recall เป็นร้อยละ 87.94, 88.28 และ 88 ตามลำดับ

ตารางที่ 2 เปรียบเทียบการวัดประสิทธิภาพของแบบจำลอง

เทคนิค	Accuracy (%)	Precision (%)	Recall (%)
Naïve Bayes	73.27	78.76	73.72
Decision Tree	87.94	88.28	88
Deep Learning	62.54	76.82	62.97
Random Forest	90.2	90.47	90.24

ส่วนแบบจำลองจาก Naïve Bayes และ Deep Learning ไม่ควรนำมาพิจารณาประยุกต์ใช้งาน เนื่องจากมีค่ามาตรวัดคุณภาพ Accuracy, Precision และ Recall ต่ำกว่าร้อยละ 80

6. สรุป

การทำวิจัยครั้งนี้มีวัตถุประสงค์เพื่อทดลองสร้างแบบจำลองทำนายแบบอากาศยานโดยใช้ข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านร่วมกับเทคนิคการทำเหมืองข้อมูล ได้แก่ Naïve Bayes, Deep Learning, Decision Tree และ Random Forest เพื่อใช้ตรวจสอบความถูกต้องของการวิเคราะห์และระบุแบบอากาศยานจากข้อมูลคุณลักษณะพื้นฐานที่ได้จากเรดาร์ รวมถึงเปรียบเทียบประสิทธิภาพความแม่นยำในการทำนาย และการใช้เวลาในการประมวลผล โดยผลการทดลองมีเพียง 2 แบบจำลองที่มีค่าความถูกต้องในการทำนายผลมากกว่าร้อยละ 80 คือ Random Forest และ Decision Tree โดยพิจารณาจากค่ามาตรวัดคุณภาพ Accuracy, Precision และ Recall ซึ่งจากผลการทดลองคณะผู้วิจัยจึงขอเสนอให้ใช้แบบจำลองจากเทคนิค Random Forest ในการประยุกต์ใช้เพื่อพัฒนากระบวนการวิเคราะห์เป้าหมายไม่ทราบฝ่ายต่อไป

7. ข้อเสนอแนะ

การวิจัยนี้เป็นการทดลองสร้างแบบจำลองทำนายแบบอากาศยานโดยใช้ข้อมูลจากระบบวิเคราะห์ความเคลื่อนไหวกำลังทางอากาศประเทศรอบบ้านของกรมข่าวทหารอากาศสำหรับใช้ในการปรับปรุงกระบวนการวิเคราะห์เป้าหมายไม่ทราบฝ่าย อย่างไรก็ตามเนื่องจากข้อมูลดังแสดงในตารางที่ 1 ยังคงมีความไม่สมดุล (Imbalance) ทำให้การทดลองจำเป็นต้องทำการ Under Sampling ข้อมูลในคลาสของแบบอากาศยานประเภทขับไล่โจมตี (Fighter), อากาศยานประเภทลำเลียง (Transporter), เฮลิคอปเตอร์ (Helicopter) ลงในจำนวนที่อาจจะมากเกินไป ซึ่งอาจจะส่งผลกระทบต่อประสิทธิภาพในการทำนายผลของแบบจำลอง ดังนั้นเพื่อเป็นการเพิ่มประสิทธิภาพการทำนายผล จึงมีความจำเป็นอย่างยิ่งที่จะต้องสำรวจข้อมูลในส่วนของคุณลักษณะอากาศยานสำหรับการฝึก (Training) เพิ่มเติมให้ครอบคลุมมากขึ้น เพื่อใช้สำหรับเป็นชุดข้อมูลสำหรับสร้างแบบจำลอง

8. เอกสารอ้างอิง

- [1] ศุภามณ จันทรสกุล. (2561). เทคนิคเหมืองข้อมูลในการวิเคราะห์ข้อมูลทางการพยาบาล. *วารสารวิชาการมหาวิทยาลัยอีสเทิร์นเอเซียฉบับวิทยาศาสตร์และเทคโนโลยี*, 12(2): 83-96.
- [2] ภัทริรา สุวรรณโค, นิสาชล จำนงศรี, จิตมนต์ อังสกุล. (2017). แบบจำลองการพยากรณ์ความเสี่ยงในการเกิดอุบัติเหตุทางถนนในเทศบาลปีใหม่ด้วยการทำเหมืองข้อมูล. *Journal of Information Science and Technology*, 7(2): 10-19.
- [3] สุวัชร ศรีเปารยะ, สายชล สิ้นสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรังกรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. *วารสารวิทยาศาสตร์และเทคโนโลยีสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง*, 25(5) : 841-853
- [4] Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic Flow Prediction With Big Data : A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2): 865-873.
- [5] บรรจบ คลกุล, จาริ ทองคำ, วาทีนิ สุขมาก. (2014). การสร้างแบบจำลองเพื่อพยากรณ์การเกิดแผลที่เท้าของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคเหมืองข้อมูล. *Journal of Science and Technology Mahasarakham University*, 33(6): 703-710.
- [6] ประเมษฐ์ ชันวานนท์, ชัยกร ยิ่งเสรี, วรพล พงษ์เพ็ชร และ ธนภัทร มังคะจิตร. (2017). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *Journal of Information Science and Technology*, 7(1): 12-21.
- [7] วิชวินท์ แสงมณี, วีระวุฒ รัตนเจริญเลิศ, ณัฐภพ โพธิรัตน์ และ เพ็ญฤทัย หนูสวัสดิ์. (2017). การสร้างโมเดลทำนายโอกาสการกลับมารักษาดูแลซ้ำของผู้ป่วยโรคเบาหวานโดยใช้เทคนิคการทำเหมืองข้อมูล. *The 5th ASEAN Undergraduate Conference in Computing (AUC2)*: 13-20.