



## การสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติโดยใช้เทคนิคเหมืองข้อมูล และการวิเคราะห์ทางสถิติ

สุนันชณา ภูแก้วงาม<sup>1\*</sup>, ณัฐธิดา หักขุนทด<sup>1</sup>, ธนพร ชันธกมล<sup>1</sup> และ สุปราณี ลิสวัสดิ์<sup>1</sup>

<sup>1</sup>สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

\*sunanchana.puk@dome.tu.ac.th

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อการเข้ามาทำงานของแรงงานข้ามชาติในประเทศไทยและเพื่อสร้างตัวแบบทางสถิติที่เหมาะสมสำหรับพยากรณ์จำนวนแรงงานข้ามชาติในประเทศไทย ด้วยเทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression) โดยใช้ข้อมูลจากสำนักงานบริหารแรงงานต่างด้าว ตั้งแต่ ปี พ.ศ. 2554 – 2564 จำนวน 132 เดือน และมีจำนวนปัจจัยที่เกี่ยวข้องทั้งหมด 9 ปัจจัย ได้แก่ ค่าแรงขั้นต่ำ อัตราการว่างงาน มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ ดัชนีค่าเงินบาท อัตราเงินเฟ้อ มูลค่าการส่งออก มูลค่าการนำเข้า จำนวนแรงงานไทยที่ได้รับอนุญาตเดินทางไปทำงานต่างประเทศ และจำนวนแรงงานข้ามชาติย้อนหลัง 1 หน่วยเวลา โดยขั้นตอนในการทำวิจัยจะดำเนินการตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (CRISP-DM) 5 ขั้นตอน ในการวิเคราะห์ปัจจัยที่เกี่ยวข้องจะใช้โปรแกรมสำเร็จรูปทางสถิติ และทดสอบประสิทธิภาพของตัวแบบโดยใช้โปรแกรม WEKA ใช้เกณฑ์ในการวัดประสิทธิภาพ 3 เกณฑ์ คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) ค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error : MAE) และค่าสัมประสิทธิ์การตัดสินใจพหุคูณ ( $R^2$ ) ผลการวิจัยพบว่าปัจจัยที่มีผลต่อการเข้ามาทำงานของแรงงานข้ามชาติในประเทศไทย คือ ค่าแรงขั้นต่ำ อัตราเงินเฟ้อ อัตราการว่างงาน มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ และจำนวนแรงงานข้ามชาติย้อนหลัง 1 หน่วยเวลา ในส่วนของการวัดประสิทธิภาพของตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติพบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีค่า RMSE เท่ากับ 15.7566 และ MAE เท่ากับ 7.7770 ซึ่งน้อยที่สุด และค่า  $R^2$  เท่ากับ 0.8847 มากเป็นอันดับ 2 รองจากเทคนิคการวิเคราะห์การถดถอยพหุคูณ ดังนั้นจึงสรุปได้ว่าตัวแบบพยากรณ์ของเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีประสิทธิภาพมากที่สุด และเหมาะสมสำหรับนำไปใช้ในการพยากรณ์จำนวนแรงงานข้ามชาติในประเทศไทย

**คำสำคัญ:** เหมืองข้อมูล การวิเคราะห์การถดถอยพหุคูณ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน แรงงานข้ามชาติ

## Model Prediction for the Number of Foreign Workers using Data Mining Techniques and Statistical Analysis

Sunanchana pukaewngam<sup>1\*</sup>, Natthatida Huekkunthod<sup>1</sup>, Thanaporn Khanthakamon<sup>1</sup>  
and Supranee Lisawadi<sup>1</sup>

<sup>1</sup>Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University

\*sunanchana.puk@dome.tu.ac.th

### Abstract

This research aims to study the factors that affect the entry of foreign workers in Thailand and to create an appropriate statistical model for forecasting the number of foreign workers in Thailand. With multiple linear regression, decision tree, and support vector regression methods using information from the foreign workers in Prof. 2011 – 2021 for 132 months, there are a total of 9 related factors, including the minimum wage, the unemployment rate, the gross domestic product value, the effective exchange rate, the inflation rate, the export value, the import value, the number of Thai workers authorized to work abroad and the number of foreign workers in the past unit of time. The research process is carried out in accordance with the 5 steps standard procedures for data mining (CRISP-DM). To analyze related factors, the statistical package is used, and the effectiveness of the model will be tested using the WEKA program. There are 3 criteria for measuring the performance of the prediction: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Multiple Determination ( $R^2$ ). The results showed that factors affecting the entry of foreign workers in Thailand are the minimum wages, the inflation rate, the unemployment rates, the gross domestic product value and the number of foreign workers in the past unit of time. The performance measurement of the model found that the support vector regression methods had the lowest MAE and RMSE, 15.7566 and 7.7770, respectively, and had  $R^2$  of 0.8847, which is the 2<sup>nd</sup> largest after multiple linear regression. Therefore, support vector regression technique is the most effective and most suitable for forecasting the number of foreign workers in Thailand.

**Keywords:** Data Mining, Multiple Linear Regression, Decision Tree, Support Vector Machine and Foreign Workers

### 1. บทนำ

สถานการณ์ตลาดแรงงานของประเทศไทยในปัจจุบันมีความต้องการแรงงานในปริมาณที่สูง เนื่องด้วยบริบทโลกที่เปลี่ยนแปลงไปและปัญหาเชิงโครงสร้างที่ตลาดแรงงานไทยกำลังเผชิญ ไม่ว่าจะเป็นการเคลื่อนย้ายแรงงานจากภาคเกษตรกรรมเข้าสู่ภาคการผลิตและภาคการบริการที่มากขึ้น การนำระบบอัตโนมัติเข้ามาทดแทนแรงงานในภาคการผลิต อัตราการเกิดที่ลดลงและการก้าวเข้าสู่สังคมผู้สูงอายุ [5] จากผลการศึกษาของสถาบันวิจัยเศรษฐกิจป๋วย อึ๊งภากรณ์ ในปี 2563 พบว่าตลาดแรงงานไทยพึ่งพาแรงงานข้ามชาติทักษะต่ำเพิ่มขึ้นมากโดยเฉพาะงานในประเภท 3D คือ งานหนัก (Difficult Job) งานสกปรก (Dirty Job) และงานอันตราย (Dangerous Job) ซึ่งเป็นงานที่คนไทยไม่นิยมทำและมีค่าแรงไม่สูงมากนัก [5] เช่น งานด้านเกษตรกรรม ประมง ก่อสร้าง เป็นต้น และพบว่าแรงงานข้ามชาติทักษะต่ำมีแนวโน้มเข้ามาทำงานในหลายประเภทธุรกิจมากขึ้น จากเดิมในปี 2553 ที่มีการจ้างแรงงานข้ามชาติเพียงไม่กี่ประเภทธุรกิจ แต่ในปี 2562 มีการจ้างแรงงานข้ามชาติในสัดส่วนที่เพิ่มสูงขึ้นมากในเกือบทุกประเภทธุรกิจ เมื่อพิจารณาการพึ่งพาแรงงานข้ามชาติของไทยเทียบกับ



ประเทศอื่น ๆ ประเทศไทยเป็นประเทศที่มีการไหลเข้าของแรงงานข้ามชาติติดอันดับที่ 17 ของโลกและยังเป็นอันดับหนึ่งในกลุ่มประเทศอาเซียน [8]

การที่แรงงานข้ามชาติเข้ามาทำงานในประเทศนั้นย่อมส่งทั้งผลดีและผลเสียต่อประเทศไทย และระบบเศรษฐกิจของไทย โดยผลดี คือ ช่วยเพิ่มมูลค่าของผลิตภัณฑ์มวลรวมระหว่างประเทศ ช่วยรักษาความได้เปรียบในการแข่งขันระยะสั้นจากการใช้แรงงานราคาถูก [8] และยังช่วยลดปัญหาการขาดแคลนแรงงานบางประเภทที่คนไทยไม่ทำ เป็นต้น แต่ถึงแม้ว่าแรงงานข้ามชาติจะส่งผลดีต่อเศรษฐกิจของไทย แต่ก็ยังมีผลกระทบในหลายๆด้านที่ตามมา เช่น ทำให้ค่าจ้างเฉลี่ยลดลง แรงงานไทยบางส่วนอาจได้รับผลกระทบจากการที่แรงงานข้ามชาติมาแย่งงานคนไทยทำ ปัญหาอาชญากรรม ปัญหาด้านยาเสพติด เป็นต้น

จากที่กล่าวมาข้างต้น ทำให้ทราบว่าแรงงานข้ามชาติมีบทบาทสำคัญต่อเศรษฐกิจไทย แต่หากจำนวนแรงงานเพิ่มขึ้นอย่างรวดเร็วและไม่มีการบริหารจัดการที่เหมาะสมก็อาจทำให้เกิดปัญหาต่าง ๆ ตามมาได้ ทางผู้วิจัยจึงสนใจที่จะศึกษาปัจจัยที่ส่งผลให้แรงงานข้ามชาติเข้ามาทำงานในประเทศไทยและสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติ โดยอาศัยเทคนิคการทำเหมืองข้อมูล (Data Mining) และการวิเคราะห์ทางสถิติ และนำตัวแบบพยากรณ์ที่ได้ไปใช้ในการแก้ไขปัญหาแรงงานข้ามชาติ วางแผนนโยบายในด้านการบริหารจัดการแรงงานข้ามชาติให้เหมาะสมและดียิ่งขึ้นในอนาคต

## 2. งานวิจัยที่เกี่ยวข้อง

ในการศึกษาการสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติโดยใช้เทคนิคเหมืองข้อมูลและการวิเคราะห์ทางสถิติ ผู้วิจัยได้ศึกษาค้นคว้า และรวบรวมรายงานวิจัยที่เกี่ยวข้องดังนี้

พิชญากร เลค (2565) งานวิจัยนี้มีวัตถุประสงค์ เปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องจักร เพื่อทำนายการเกิดอุบัติเหตุ โดยใช้การวิเคราะห์อนุกรมเวลา ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ได้แก่ การถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลในจังหวัดที่มีจำนวนการเกิดอุบัติเหตุมากที่สุด ได้แก่ จังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ โดยรวบรวมข้อมูลตั้งแต่ปี พ.ศ. 2558-2562 จำนวน 60 เดือน จากการทดลองพบว่าแบบจำลองที่มีความเหมาะสมในการพยากรณ์การเกิดอุบัติเหตุในประเทศไทย ดังต่อไปนี้แบบ จำลองการถดถอยเชิงเส้น มีความเหมาะสมมากที่สุดกับชุดข้อมูลจังหวัดกรุงเทพมหานคร จังหวัดนครราชสีมา และจังหวัดสมุทรปราการ มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับร้อยละ 16.29 19.29 และร้อยละ 19.64

วีรศักดิ์ ฟองเงิน, วรภา อารีราษฎร์ และเผด็จ พรหมสาขา ณ สกลนคร (2560) ได้ศึกษาการพยากรณ์ปริมาณน้ำในเขื่อนก๊วลม โดยใช้เทคนิคเหมืองข้อมูล การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาเทคนิคเหมืองข้อมูลที่เหมาะสมในการพยากรณ์ปริมาณน้ำในเขื่อน และ 2) เปรียบเทียบผลปริมาณน้ำจริงกับผลการพยากรณ์ปริมาณน้ำในเขื่อนก๊วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้ได้บริหารจัดการข้อมูลน้ำที่มีอยู่ที่เป็นปัจจัยต่อการเปลี่ยนแปลงระดับน้ำ ประกอบด้วยปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในเขื่อน ปริมาณการปล่อยน้ำ และอัตราการระเหย โดยรวบรวมข้อมูลรายวันตั้งแต่ปี พ.ศ. 2535 – พ.ศ. 2559 รวม 25 ปี จำนวน 9,125 ระเบียบ เพื่อนำมาพยากรณ์ด้วยเทคนิคการพยากรณ์ผลการวิจัยพบว่า 1) เทคนิคเหมืองข้อมูลที่นำมาใช้ในการพยากรณ์ปริมาณน้ำในเขื่อนประกอบด้วย 4 เทคนิค ได้แก่ เทคนิควิธีการวิเคราะห์การถดถอย วิธีโครงข่ายประสาทเทียม วิธีแบบจำลองต้นไม้เอ็มไพร์พีและวิธี เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยเลือกจากงานวิจัยที่เกี่ยวข้องที่มีการใช้งานมากที่สุดและ 2) ผลการ เปรียบเทียบการพยากรณ์ปริมาณน้ำรายเดือนในเขื่อนก๊วลม จังหวัดลำปาง โดยใช้เทคนิคเหมืองข้อมูลทั้ง 4 เทคนิค พบว่า วิธีแบบจำลองต้นไม้เอ็มไพร์พี มีค่าสัมบูรณ์ของความคลาดเคลื่อนต่ำสุดที่ 10.58 และเป็นวิธีที่เหมาะสมที่สุดสำหรับนำไปพัฒนาระบบเพื่อการพยากรณ์น้ำในเขื่อน ทั้งนี้เมื่อพิจารณาค่าสัมบูรณ์ของความคลาดเคลื่อนแต่ละเทคนิคเรียงตามค่าความคลาดเคลื่อนจากน้อยไปมาก พบว่า วิธีแบบจำลองต้นไม้เอ็มไพร์พีมีค่าเท่ากับ 10.58 วิธีซัพพอร์ตเวกเตอร์แมชชีน มีค่าเท่ากับ 19.90 วิธีวิเคราะห์การถดถอย มีค่าเท่ากับ 20.00 และวิธีโครงข่ายประสาทเทียม มีค่าเท่ากับ 21.65 ตามลำดับ

Navin and G. Vadivu (2015) งานวิจัยฉบับนี้ได้พัฒนาโมเดลพยากรณ์ราคาทองคำโดยพิจารณาจากปัจจัยทางเศรษฐกิจ เช่น อัตราเงินเฟ้อ การเคลื่อนไหวของราคาเงินตรา และปัจจัยอื่นๆ โดยใช้วิธีที่เหมาะสมในการพยากรณ์ คือ วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย และวิธีต้นไม้ตัดสินใจ จากการทดลองได้ว่ามี 5 ปัจจัยที่มีผลต่อราคาทองคำ ได้แก่ ค่าเปิด ค่าปิด ค่าต่ำสุด ค่าสูงสุด และปริมาณการซื้อขาย ผลของทั้งสองวิธีคือ วิธีต้นไม้ตัดสินใจดีที่สุดสำหรับการเลือกปัจจัย และวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยใช้ได้ดีที่สุดกับชุดข้อมูลที่มีขนาดใหญ่ แต่จะใช้เวลานานในการทดสอบ

อย่างไรก็ตามวิธีต้นไม้ตัดสินใจจะใช้เวลาในการประมวลผลและให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) น้อยกว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย

Abdul-Lateef Balogun (2022) ได้ศึกษาผลกระทบของตัวแปรด้านภูมิอากาศที่มีผลต่อความเข้มข้นของโอโซนในมาเลเซีย เพื่อให้เข้าใจถึงความเชื่อมโยงระหว่างการเปลี่ยนแปลงสภาพภูมิอากาศและความเข้มข้นของโอโซน ในการวิเคราะห์ตัวแปรด้านภูมิอากาศและความเข้มข้นของโอโซนจะใช้การวิเคราะห์ความสัมพันธ์ (Linear Regression) และเทคนิคเหมืองข้อมูล ซึ่งได้แก่ เทคนิคการสุ่มป่าไม้ (Random Forest), เทคนิคต้นไม้ตัดสินใจ (Decision Tree), เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression) ผลการวิจัยพบว่าตัวแปรด้านภูมิอากาศที่มีผลต่อความเข้มข้นของโอโซน คือ อุณหภูมิและความเร็วลมซึ่งมีความสัมพันธ์กันในเชิงลบ และการตรวจสอบความแม่นยำของตัวแปรด้านภูมิอากาศพบว่า เทคนิคการสุ่มป่าไม้ให้ค่าความสัมพันธ์การตัดสินใจพหุคูณมากที่สุด เท่ากับ 0.970 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด เท่ากับ 2.737 และค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อน เท่ากับ 1.824 ตามด้วยการวิเคราะห์ความสัมพันธ์ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย และเทคนิคต้นไม้ตัดสินใจตามลำดับ การวิจัยฉบับนี้จัดทำขึ้นเพื่อเป็นแนวทางให้รัฐบาลมาเลเซียรับมือกับปัญหาด้านสภาพอากาศเพื่อลดการเปลี่ยนแปลงของสภาพอากาศที่เป็นมลพิษและมีการพัฒนาที่ยั่งยืนในอนาคต

### 3. วิธีการดำเนินการวิจัย

การสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติได้ดำเนินการตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining หรือ CRISP-DM) ซึ่งแบ่งออกเป็น 5 ขั้นตอน ดังนี้

#### 3.1 ขั้นตอนที่ 1 การทำความเข้าใจปัญหา (Research Understanding)

จากการศึกษางานวิจัยที่เกี่ยวข้องกับแรงงานข้ามชาติและการเข้ามาทำงานของแรงงานข้ามชาติในประเทศไทย พบว่า ในช่วง 10 ปีที่ผ่านมาจำนวนแรงงานข้ามชาติเฉลี่ยมีแนวโน้มเพิ่มขึ้นในทุกๆ ปี โดยมีฤดูกาลหรือวัฏจักรของแรงงานข้ามชาติที่เข้ามาทำงานในประเทศไทยอย่างไม่แน่นอน [8] เนื่องจากสาเหตุต่างๆทั้งในประเทศไทย และประเทศต้นทางของแรงงานข้ามชาติ ส่วนสาเหตุและปัจจัยที่ทำให้แรงงานข้ามชาตินิยมเข้ามาทำงานในประเทศไทย มีหลายสาเหตุ เช่น ความแตกต่างทางเศรษฐกิจของประเทศ ค่าแรงขั้นต่ำ การขาดแคลนแรงงานของประเทศไทย สวัสดิการแรงงานของไทย เป็นต้น

#### 3.2 ขั้นตอนที่ 2 การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)

ข้อมูลที่นำมาใช้ในการสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติเป็นข้อมูลทุติยภูมิ (Secondary Data) และเป็นข้อมูลรายเดือนย้อนหลัง 132 เดือน ตั้งแต่ปี พ.ศ. 2554 – 2564 รวมทั้งหมด 1,188 ระเบียบ

#### 3.3 ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

3.3.1 นำเข้าข้อมูลสำหรับการวิเคราะห์ข้อมูลดิบ โดยการวิจัยครั้งนี้ใช้ข้อมูลทุติยภูมิ (Secondary Data) เป็นข้อมูลรายเดือนย้อนหลัง 132 เดือน ตั้งแต่ปี พ.ศ. 2554 – 2564 รวมทั้งหมด 1,188 ระเบียบ

3.3.2 ตรวจสอบข้อมูลสูญหายในแต่ละตัวแปร หากพบข้อมูลสูญหายจะแทนค่าด้วยค่าเฉลี่ย

3.3.3 กำหนดและตรวจสอบความถูกต้องของประเภทข้อมูลในแต่ละตัวแปร ดังตารางที่ 1



ตารางที่ 1 รายละเอียดตัวแปรที่ใช้ในการศึกษา

ตัวแปร	คำอธิบาย	ประเภทข้อมูล	แหล่งที่มาของข้อมูล
num_of_foreign_workers	จำนวนแรงงานข้ามชาติ (คน)	จำนวนจริง	สำนักงานบริหารแรงงานต่างด้าว
num_of_foreign_workersback (X1)	จำนวนแรงงานข้ามชาติย้อนหลัง 1 หน่วยเวลา (คน)	จำนวนจริง	
minimum_wage (X2)	ค่าแรงขั้นต่ำต่อเดือน (บาท)	จำนวนจริง	หอสมุดแห่งมหาวิทยาลัยธรรมศาสตร์
Gdp (X3)	มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ (ล้านบาท)	จำนวนจริง	
effective_exchange_rate (X4)	ดัชนีค่าเงินบาท	จำนวนจริง	ธนาคารแห่งประเทศไทย
inflation_rate (X5)	อัตราเงินเฟ้อ (ร้อยละ)	จำนวนจริง	
unemployment_rate (X6)	อัตราการว่างงาน (ร้อยละ)	จำนวนจริง	
export_value (X7)	มูลค่าการส่งออก (ร้อยละ)	จำนวนจริง	สำนักงานปลัดกระทรวงพาณิชย์
import_value (X8)	มูลค่าการนำเข้า (ร้อยละ)	จำนวนจริง	
num_of_Thais_working_abroad (X8)	จำนวนคนไทยที่ไปทำงานที่ต่างประเทศ (คน)	จำนวนจริง	กองบริหารแรงงานไทยไปต่างประเทศ

ตารางที่ 2 ตัวอย่างข้อมูลที่ใช้ในการศึกษา 12 เดือนแรก

เดือน	X1	X2	X3	X4	X5	X6	X7	X8	X9	ตัวแปรตาม
1	5590	0.98	2960302	151.7	29.92	30.31	1.8492	3.81	133.5155	134.5728
2	5590	0.7	2960302	153.8	30.37	30.74	1.8492	3.81	134.57	76.554
3	5590	0.72	2960302	147.9	30.4	30.78	1.8492	3.81	76.55	91.4525
4	5590	0.75	2810921	128.9	30.04	30.42	2.3955	3.81	91.45	91.5028
5	5590	0.53	2810921	123.2	29.8	30.18	2.3955	3.81	91.5	91.4644
6	5590	0.42	2810921	123.2	30.07	30.44	2.3955	3.81	91.46	99.6142
7	5590	0.52	2852152	135.7	30.3	30.68	2.5967	3.81	99.61	100.621
8	5590	0.68	2852152	132.7	29.7	30.08	2.5967	3.81	100.62	169.0036
9	5590	0.75	2852152	124.8	29.62	30.01	2.5967	3.81	169	181.047
10	5590	0.56	2683531	148.6	30.26	30.64	2.1823	3.81	181.05	186.1639
เดือน	X1	X2	X3	X4	X5	X6	X7	X8	X9	ตัวแปรตาม
11	5590	0.82	2683531	129	30.59	30.97	2.1823	3.81	186.16	196.2858
12	5590	0.43	2683531	129.2	30.82	31.2	2.1823	3.81	196.29	195.065

3.3.4 กำหนดหน้าที่ของตัวแปรโดยกำหนดให้จำนวนแรงงานข้ามชาติเป็นตัวแปรตาม (Dependent Variable) สำหรับการสร้างตัวแบบพยากรณ์ ซึ่งเป็นข้อมูลเชิงปริมาณมีหน่วยเป็นคน

3.3.5 ทดสอบระดับความสัมพันธ์ของตัวแปร เพื่อให้ตัวแปรที่เลือกใช้มีความเหมาะสมต่อตัวแบบพยากรณ์ ทำการทดสอบค่าความสัมพันธ์ระหว่างตัวแปรอิสระทั้งหมดกับตัวแปรตาม ภายใต้ระดับความเชื่อมั่น 95% โดยได้ทำการทดสอบสหสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามโดยใช้ Pearson's Correlation Coefficient ตามสมมติฐานทางสถิติ ที่ระดับนัยสำคัญ 0.05 ดังนี้

$H_0$  : ตัวแปรสองตัวไม่มีความสัมพันธ์กัน (ไม่มีสหสัมพันธ์กัน)

$H_1$  : ตัวแปรสองตัวมีความสัมพันธ์กัน (มีสหสัมพันธ์กัน)

การวิเคราะห์ค่าสหสัมพันธ์จะพิจารณาจากค่า P-Value หากค่า P-Value < 0.05 คือ ปฏิเสธ  $H_0$  ตัวแปรทั้งสองตัวมีความสัมพันธ์กัน

#### 3.4 ขั้นตอนที่ 4 การสร้างแบบจำลอง (Modeling)

นำตัวแปรที่ได้คัดเลือกในขั้นตอนที่ 3 มาสร้างตัวแบบพยากรณ์ด้วย 3 เทคนิค ดังนี้

3.4.1 เทคนิคการวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) สร้างตัวแบบพยากรณ์ผ่านโปรแกรมสำเร็จรูปทางสถิติ โดยเลือกใช้วิธีการเลือกตัวแปรแบบลำดับขั้น (Stepwise Selection)

3.4.2 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยใช้อัลกอริทึมต้นไม้ (Tree) คือ Reduced Error Pruning Tree (REPTree) ผ่านโปรแกรม WEKA โดยใช้ค่าพารามิเตอร์ตามค่าตั้งต้นของโปรแกรม

3.4.3 เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression) โดยใช้ Polynomial Kernel function และกำหนดค่า Cost = 1000 ผ่านโปรแกรม WEKA

#### 3.5 ขั้นตอนที่ 5 การประเมินผล (Evaluation)

ทดสอบประสิทธิภาพของตัวแบบพยากรณ์ทั้ง 3 เทคนิค โดยวัดประสิทธิภาพด้วยวิธี Cross-Validation 10-fold หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) ค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error : MAE) และค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (Coefficient of Multiple Determination :  $R^2$ ) โดยเปรียบเทียบค่าความคลาดเคลื่อนทั้ง 3 ค่า

### 4. ผลการวิจัย

#### 4.1 การวิเคราะห์ความสัมพันธ์ของตัวแปรตามและตัวแปรอิสระ

จากปัจจัยทั้ง 9 ปัจจัย ทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามจำนวนแรงงานข้ามชาติ เพื่อคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามมาใช้ในการสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติ ซึ่งจะคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) โดยใช้โปรแกรมสำเร็จรูป SPSS เป็นเครื่องมือในการคำนวณค่าสถิติ ที่ระดับความเชื่อมั่น 95% ได้ผลดังตารางที่ 1





ตารางที่ 1 ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามจำนวนแรงงานข้ามชาติ

	num_of_foreign_workers			num_of_foreign_workers	
minimum_wage	Pearson Correlation	0.589**	unemployment_rate	PearsonCorrelation	0.475**
	Sig. (2-tailed)	0.000		Sig. (2-tailed)	0.000
gdp	Pearson Correlation	0.676**	export_value	Pearson Correlation	-0.088
	Sig. (2-tailed)	0.000		Sig. (2-tailed)	0.313
effective_exchange_rate	Pearson Correlation	-0.140	import_value	Pearson Correlation	-0.078
	Sig. (2-tailed)	-0.110		Sig. (2-tailed)	0.372
inflation_rate	Pearson Correlation	0.507**	num_of_Thais_working_abroad	Pearson Correlation	-0.016
	Sig. (2-tailed)	0.000		Sig. (2-tailed)	0.854
num_of_foreign_workersback	Pearson Correlation	-0.499**			
	Sig. (2-tailed)	0.000			

\*\* .correlation is significant at the 0.01

\* .correlation is significant at the 0.05

จากตารางแสดงค่าสหสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามจำนวนแรงงานข้ามชาติ สรุปได้ว่าตัวแปรอิสระที่มีความสัมพันธ์กันอย่างมีนัยสำคัญกับจำนวนแรงงานข้ามชาติ คือ ค่าแรงขั้นต่ำต่อเดือน มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) อัตราการว่างงาน อัตราเงินเฟ้อ และจำนวนแรงงานข้ามชาติย้อนหลังไป 1 หน่วยเวลา จึงเลือกใช้ตัวแปรเหล่านี้เป็นตัวแปรอิสระเพื่อใช้ในการสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติ โดยให้เป็น  $X_1, X_2, X_3, X_4$  และ  $X_5$  ตามลำดับ

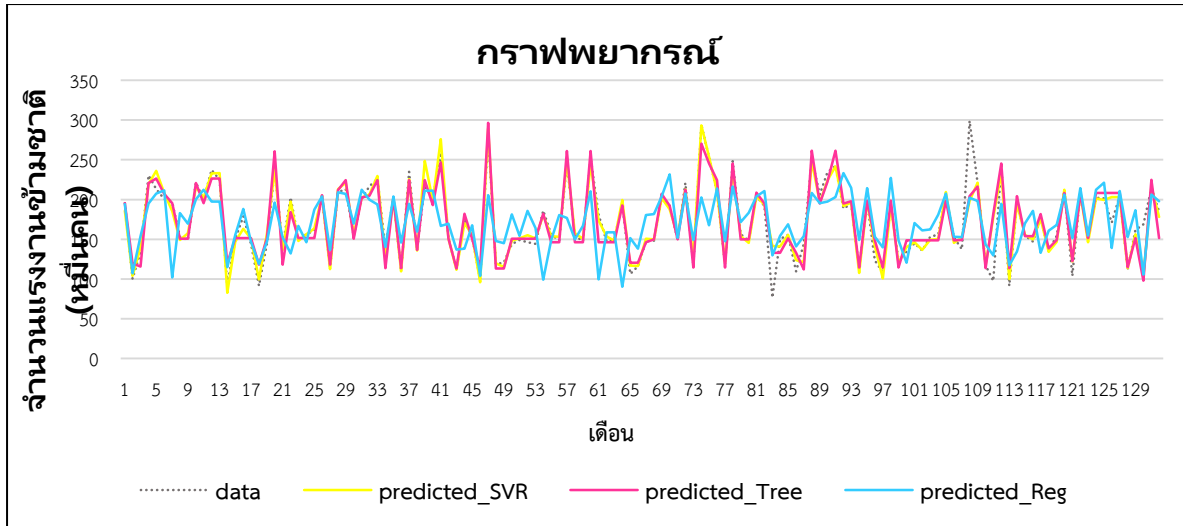
#### 4.2 การทดสอบประสิทธิภาพของตัวแบบพยากรณ์ทั้ง 3 เทคนิค

จากนั้นนำตัวแปรที่ทำการคัดเลือกไปสร้างตัวแบบพยากรณ์ด้วย 3 เทคนิค คือ การวิเคราะห์การถดถอยพหุคูณ (Multiple Linear Regression) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression) และวัดประสิทธิภาพด้วยวิธี Cross-Validation 10-folds คือ การแบ่งข้อมูลออกเป็น 10 ส่วน โดยที่แต่ละส่วนมีข้อมูลเท่ากัน หลังจากนั้นข้อมูลส่วนหนึ่งจะใช้เป็นตัวทดสอบประสิทธิภาพของตัวแบบพยากรณ์เพื่อทำการวิเคราะห์ทดสอบข้อมูลหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) ค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error : MAE) และค่าสัมประสิทธิ์การตัดสินใจพหุคูณ ( $R^2$ ) ซึ่งจะเปรียบเทียบค่าความคลาดเคลื่อนทั้ง 3 ค่า

ตารางที่ 2 ผลการทดสอบประสิทธิภาพของตัวแบบพยากรณ์ทั้ง 3 เทคนิค

ประเภทของตัวแบบ	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE)	ค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อน (MAE)	ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ ( $R^2$ )
1. Multiple Linear Regression	16.1456	8.2543	0.8920
2. Decision Tree	17.9331	10.5824	0.8527
3. Support Vector Regression	15.8308	7.7328	0.8847

จากผลการทดสอบประสิทธิภาพของตัวแบบพยากรณ์ทั้ง 3 เทคนิค จะเห็นได้ว่าตัวแบบพยากรณ์ของเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีประสิทธิภาพดีที่สุด เนื่องจากมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อนต่ำที่สุด เท่ากับ 15.8308 และ 7.7328 ตามลำดับ นอกจากนี้ยังให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณเท่ากับ 0.8847 มากเป็นอันดับสองรองจากเทคนิคการวิเคราะห์การถดถอยพหุคูณ



ภาพที่ 1 กราฟเปรียบเทียบค่าจริงของจำนวนแรงงานข้ามชาติและค่าพยากรณ์จากเทคนิคทั้ง 3 เทคนิค

เมื่อพิจารณาภาพที่ 1 จะเห็นได้ว่าตัวแบบพยากรณ์ของเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีค่าพยากรณ์ใกล้เคียงกับค่าจริงมากที่สุด ซึ่งสอดคล้องกับผลการทดสอบประสิทธิภาพข้างต้น ดังนั้นตัวแบบพยากรณ์ของเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยจึงเหมาะสำหรับการนำไปใช้ในการสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติมากที่สุด

## 5. อภิปรายผลการวิจัย

การสร้างตัวแบบพยากรณ์จำนวนแรงงานข้ามชาติโดยใช้เทคนิคเหมืองข้อมูลและการวิเคราะห์ทางสถิติ เพื่อศึกษาปัจจัยที่มีผลต่อการเข้ามาทำงานของแรงงานข้ามชาติในประเทศไทย หากพิจารณาแยกทั้ง 3 เทคนิค คือ เทคนิคการวิเคราะห์การถดถอยพหุคูณ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression) ได้ผลดังนี้ เทคนิคการวิเคราะห์การถดถอยพหุคูณ ปัจจัยที่มีผลต่อจำนวนแรงงานข้ามชาติในประเทศไทย ได้แก่ มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) และจำนวนแรงงานข้ามชาติย้อนหลังไป 1 หน่วยเวลา เทคนิคต้นไม้ตัดสินใจ ปัจจัยที่มีผลมากที่สุด คือ อัตราเงินเฟ้อ มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) และจำนวนแรงงานข้ามชาติย้อนหลังไป 1 หน่วยเวลา ตามลำดับ และเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยนั้น ปัจจัยที่มีผลต่อการเข้ามาทำงานของแรงงานข้ามชาติ คือ ค่าแรงขั้นต่ำ อัตราเงินเฟ้อ อัตราการว่างงาน มูลค่าผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) และจำนวนแรงงานข้ามชาติย้อนหลังไป 1 หน่วยเวลา เนื่องจากในแต่ละเทคนิคมีจำนวนปัจจัยที่แตกต่างกัน อาจส่งผลต่อประสิทธิภาพของตัวแบบพยากรณ์ได้ เทคนิคที่มีจำนวนปัจจัยมากกว่าจะได้เปรียบและมีประสิทธิภาพมากกว่าเทคนิคที่มีจำนวนปัจจัยน้อยกว่า ในส่วนของการศึกษาการสร้างตัวแบบพยากรณ์พยากรณ์จำนวนแรงงานข้ามชาติในประเทศไทย ทำการสร้างตัวแบบพยากรณ์ด้วย 3 เทคนิค จากนั้นนำมาเปรียบเทียบประสิทธิภาพของเทคนิคการพยากรณ์ โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) ค่าเฉลี่ยสมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error : MAE) และค่าสัมประสิทธิ์การตัดสินใจพหุคูณ ( $R^2$ ) ได้ผลดังนี้คือ เทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีประสิทธิภาพดีที่สุดในเรื่องมีค่า MAE = 7.7770 และ RMSE = 15.7566 น้อยที่สุด อีกทั้งเทคนิคนี้ยังมีจำนวนปัจจัยมากที่สุดอีกด้วย จึงสรุปได้ว่าตัวแบบพยากรณ์ของเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีประสิทธิภาพดีที่สุด และเหมาะสำหรับนำไปพยากรณ์จำนวนแรงงานข้ามชาติในประเทศไทย





## 6. เอกสารอ้างอิง

- [1] Aphinya Kangthong. (2021). *Forecasting Plastic Resin consumption in Thailand*. [Master of Business Administration]. Thammasat University. TU Digital Collection.  
[https://digital.library.tu.ac.th/tu\\_dc/frontend/Info/item/dc:270732](https://digital.library.tu.ac.th/tu_dc/frontend/Info/item/dc:270732). (In Thai)
- [2] Chanin Keawwiboonphan. (2021). *Model Averaging Over Regression Trees for Prediction*. [Master of Science (Applied Statistics)]. Thammasat University. TU Digital Collection.  
[https://digital.library.tu.ac.th/tu\\_dc/frontend/Info/item/dc:271629](https://digital.library.tu.ac.th/tu_dc/frontend/Info/item/dc:271629) (In Thai)
- [3] Chanita Hangkaew, Juthaphorn Sinsomboonthong, and Thidaporn Supapakorn. (2020). Prediction of choosing types of insurance for the customers. *Huachiew Chalermprakiet Science and Technology Journal*, 6(2), 81-95. (In Thai)
- [4] Chitpong Kittinaradorn. (2020). *Support Vector Machines*. Machine Learning and Full-stack Python.  
<https://guopai.github.io/index.html> (In Thai)
- [5] Chommanart Nittayo, Sophon Tunyavetchakit, Bovonvich Jindarak, and Nuntanid Thongsri. (17 June 2020). Facts about Foreign Workers in Thailand: Part 1 Low-skilled workers. Puey Ungphakorn Institute for Economics Research. <https://www.pier.or.th/en/abridged/2020/14/> (In Thai)
- [6] Jaruek Khongnantha. (2021). *Forecasting of Solar Rooftop Consumption in Thailand*. [Master of Business Administration]. Thammasat University. TU Digital Collection.  
[https://digital.library.tu.ac.th/tu\\_dc/frontend/Info/item/dc:270738](https://digital.library.tu.ac.th/tu_dc/frontend/Info/item/dc:270738) (In Thai)
- [7] Pitchayakorn Lake. (2022). Forecasting retail oil and natural gas vehicles prices in Thailand using time series data mining techniques. *Journal of Positive School Psychology*, 6(3), 9537-9541. (In Thai)
- [8] Navin, K., & Vadivu, G. (2015). Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector Regression (SVR). *International Journal of Science and Research (IJSR)*, 4(3), 2026-2030.
- [9] Office of Permanent Secretary for Ministry of Labour. (2021). Economic Situation Report International Labor Market and International Labor Migration Quarter 1/2021.  
<http://warning.mol.go.th/uploadFile/pdf/pdf-2022-08-29-1661761146.pdf> (In Thai)
- [10] Pajjit Suksomboon, and Nicha Naphaporn Chongkasikit. (2017). Effects of Models that Affect their Ability to Repay Loans of Member in Credit Union Cooperative. *Industrial Technology Lampang Rajabhat University Journal*, 10(1), 101-110. (In Thai)
- [11] Phaksinipitch, K., Chalermpon, J., & Manoon, T. (2019). Forecasting the Number of Migrant Workers in Thailand: Empirical Study and Discussion. *St. Theresa Journal of Humanities and Social Sciences*, 5(2), 1-19.
- [12] Ronnachai Cheunthawat, Kittsak Kerdprasop, and Nittaya Kerdprasop. (2017). The Demand Forecasting of Electrical Power Distribution Units by Support Vector Regression with 3 Folds Cross-Validation. *Journal of Science & Technology, Ubon Ratchathani University*, 19(1), 215-232. (In Thai)
- [13] Siriporn Praneechon. (2019). *Forecasting The Renewal Rate of Voluntary Motor Insurance*. [Master of Business Administration]. Thammasat University. TU Digital Collection.  
[https://digital.library.tu.ac.th/tu\\_dc/frontend/Info/item/dc:173278](https://digital.library.tu.ac.th/tu_dc/frontend/Info/item/dc:173278) (In Thai)
- [14] Thanawat Wannaprapha. (2021). *Using Data Mining Techniques for screening people successfully in undergraduate studies of Educational Technology*. [Master of Engineering]. Dhurakij Pundit University. Digital Research Information Center.  
<https://dric.nrct.go.th/Search/ShowFulltext/2/312114> (In Thai)



- [15] Thiptida Vongpipan. (2013). *Lease Approval Using Data Mining Techniques*. [Master of Science]. Dhurakij Pundit University. Digital Research Information Center. <https://dric.nrct.go.th/Search/ShowFulltext/2/274073> (In Thai)
- [16] Werasak Fongngen, Worapapha Arreerard, and Padej Phomasakha Na Sakolnakorn. (2018). Forecasting Daily Discharge in Kiewlom Dam Using Data Mining Techniques. *Journal of Technology Management Rajabhat Maha Sarakham University*, 4(1), 27-33. (In Thai)