

บทที่ 5

สรุปผลการดำเนินงาน และข้อเสนอแนะ

จากผลการวิเคราะห์การเปรียบเทียบโมเดลเพื่อพยากรณ์แรงงานไทยในต่างประเทศ ด้วยเทคนิคเหมืองข้อมูล และการจัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูล ทางคณะผู้จัดทำได้ทำการสรุปผลการทำโครงการ ข้อจำกัดของเว็บไซต์ ปัญหาและอุปสรรคต่าง ๆ ของการทำโครงการ และข้อเสนอแนะเพื่อพัฒนาโครงการ ดังนี้

5.1 บทสรุปผลโครงการ

โครงการฉบับนี้ถูกจัดทำตามหลักการวิเคราะห์ข้อมูล ผู้จัดทำได้ทำการรวบรวมเก็บข้อมูลสถิติแรงงานไทยที่เดินทางไปทำงานต่างประเทศ จากกองบริหารแรงงานไทยไปต่างประเทศ กรมการจัดหางาน ช่วงปี พ.ศ. 2564–2567 จำนวน 324,497 ระเบียบ ดำเนินการตามกระบวนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล (Data Mining) โดยใช้กระบวนการมาตรฐาน CRISP-DM โดยตั้งเป้าหมายเพื่อเปรียบเทียบประสิทธิภาพของโมเดลพยากรณ์ 4 โมเดล ได้แก่ 1) Multiple Linear Regression 2) Gradient Boosted Trees (GBT) 3) Random Forest Regression และ 4) k-Nearest Neighbors (kNN) โดยใช้กระบวนการทำการฝึกโมเดล โดยใช้ชุดข้อมูลสอน (Training Set) 70% และชุดข้อมูลทดสอบ (Testing Set) 30% เพื่อประเมินผลลัพธ์ของโมเดล และวัดประสิทธิภาพ (Performance) โดยดำเนินการผ่านโปรแกรม RapidMiner

จากการเปรียบเทียบประสิทธิภาพของทั้ง 4 โมเดลโดยใช้ค่าความผิดพลาด RMSE และ MAE พบว่า Multiple Linear Regression (MLR) เป็นโมเดลที่เหมาะสมที่สุดสำหรับชุดข้อมูลนี้ โดยให้ค่าความคลาดเคลื่อนต่ำที่สุด คือค่า RMSE เท่ากับ 67.683 และ MAE เท่ากับ 29.764 ซึ่งแสดงให้เห็นว่าแบบจำลองมีความแม่นยำและเสถียรภาพในการพยากรณ์ยอดรวมแรงงานรายเดือนได้ดีกว่าโมเดลอื่น ๆ

ทางคณะผู้จัดทำนำผลลัพธ์ที่ทำการวิเคราะห์แล้ว นำมาสร้างเป็นสารสนเทศจัดแสดงผ่านเว็บไซต์เพื่อเผยแพร่ข้อมูลสถิติและการพยากรณ์แรงงานไทยในต่างประเทศ โดยใช้ชุดคำสั่ง HTML และ CSS ร่วมกับการนำเสนอข้อมูลแบบ Visualization ที่ประกอบด้วย แดชบอร์ดสรุปสถิติ กราฟแสดงแนวโน้ม และกราฟเปรียบเทียบผลการพยากรณ์ ด้วยโปรแกรม Power BI และเผยแพร่ผลการวิเคราะห์ข้อมูลผ่านเว็บไซต์เพื่อเป็นประโยชน์ต่อผู้ที่สนใจและหน่วยงานที่เกี่ยวข้อง

5.2 ข้อจำกัดเว็บไซต์

5.2.1 ผู้ใช้ทั่วไปไม่สามารถนำเข้าหรืออัปโหลดไฟล์ข้อมูลดิบชุดใหม่เพื่อทำการพยากรณ์แบบ Real-time ได้ด้วยตนเอง

5.2.2 ระบบเว็บไซต์เน้นการแสดงผล (Visualization) จากผลลัพธ์ที่ผ่านการวิเคราะห์มาแล้ว ไม่รองรับการปรับเปลี่ยนพารามิเตอร์ของโมเดลผ่านหน้าเว็บไซต์โดยตรง

5.3 ปัญหาและอุปสรรคของโครงการ

5.3.1 ข้อมูลที่ได้รับมีความซ้ำซ้อนและเป็นข้อความ (Text) จำนวนมาก ทำให้เสียเวลาในขั้นตอนทำความสะอาดและแปลงข้อมูล (Data Preparation) นานกว่าขั้นตอนอื่น

5.3.2 การปรับจูนพารามิเตอร์โมเดลใน RapidMiner มีความซับซ้อน ต้องใช้เวลาศึกษาและทดลองปรับค่าค่อนข้างนานเพื่อให้ได้ผลลัพธ์ที่แม่นยำ

5.3.3 เดิมคณะผู้จัดทำวางแผนจะใช้โมเดลขั้นสูง (Holt-Winters, SARIMA, LSTM) แต่เนื่องจาก RapidMiner เวอร์ชันการศึกษาที่มีข้อจำกัดในการรองรับ จึงจำเป็นต้องปรับเปลี่ยนขอบเขตมาใช้โมเดล Machine Learning พื้นฐาน (เช่น Linear Regression, Random Forest) ร่วมกับเทคนิค Windowing แทน

5.4 ข้อเสนอแนะ

5.4.1 ควรนำปัจจัยทางเศรษฐกิจอื่น ๆ เข้ามาร่วมวิเคราะห์ เพื่อให้โมเดลสามารถอธิบายความผันผวนของข้อมูลได้ดียิ่งขึ้น

5.4.2 ในอนาคตควรเปลี่ยนไปใช้การเขียนโปรแกรม (เช่น Python) แทน RapidMiner เพื่อให้มีความยืดหยุ่นในการสร้างโมเดลเชิงลึก (Deep Learning) อย่าง LSTM หรือ SARIMA ซึ่งเหมาะสมกับข้อมูลอนุกรมเวลามากกว่า