บทที่ 4

ผลการดำเนินงาน

ผลการดำเนินงานโครงการเรื่องการพัฒนาแบบจำลองการพยากรณ์แรงงานไทยในต่างประเทศด้วยเทคนิคเหมืองข้อมูล (Data Mining) และใช้กระบวนการวิเคราะห์ข้อมูลด้วยกระบวนการ CRISP-DM ที่สำคัญหลากหลายขั้นตอน เมื่อเสร็จสิ้นจากกระบวนการวิเคราะห์ข้อมูลแล้วจะเป็นขั้นตอนการสร้างโมเดลจำแนกประเภทข้อมูลในขั้นตอน Modeling โดยจะใช้เทคนิคดังต่อไปนี้ 1) Multiple Linear Regression 2) Gradient Boosted Trees (GBT) 3) Random Forest Regression และ 4) k-Nearest Neighbors (kNN) ซึ่งจะนำมาเปรียบเทียบความแม่นยำของแบบจำลองทั้งหมด โดยจะใช้ตัวชี้วัดมาตรฐาน ดังนี้ RMSE (Root Mean Squared Error) และ MAE (Mean Absolute Error) เป็นตัวเปรียบเทียบ จากนั้นทำการนำโมเดลที่ได้ไปทำการพัฒนาเว็บไซต์สำหรับการทำนายแสดงผล

4.1 ผลการดำเนินงานการวิเคราะห์ข้อมูลด้วย CRISP-DM

4.1.1 ผลการทำความเข้าใจปัญหา (Problem Understanding Results) พบว่าข้อมูลจำนวนแรงงานไทยที่เดินทางไปทำงานต่างประเทศมีความผันผวนตามช่วงเวลา (Seasonality) และได้รับอิทธิพลจากปัจจัยภายนอกหลายประการ ทำให้ยากต่อการวางแผนบริหารจัดการด้วยการคาดคะเน ผลจากการศึกษานี้ได้ข้อสรุปที่ชัดเจนว่า เป้าหมายหลักของโครงการคือการพยากรณ์ "ยอดรวม (Total)" ของแรงงานรายเดือน เพื่อให้หน่วยงานภาครัฐสามารถจัดเตรียมทรัพยากรและกำหนดนโยบายรองรับได้อย่างแม่นยำ

4.1.2 ผลการทำความเข้าใจข้อมูล (Data Understanding Results) จากการรวบรวมและสำรวจข้อมูลดิบ (Raw Data) จากกองบริหารแรงงานไทยไปต่างประเทศ ช่วงปี พ.ศ. 2564-2567 จำนวนรวม 324,497 ระเบียน พบลักษณะสำคัญของข้อมูล ดังนี้

ลักษณะข้อมูล ข้อมูลดิบเป็นรายระเบียน (Transactional Data) ที่ประกอบด้วยรายละเอียดรายบุคคล เช่น เพศ อายุ จังหวัด และวิธีการเดินทาง ซึ่งยังไม่สามารถนำมาพยากรณ์ภาพรวมได้ทันที จำเป็นต้องมีการแปลงสภาพข้อมูล

คุณภาพข้อมูล พบปัญหาข้อมูลเชิงข้อความ (Text) มีความซ้ำซ้อนในบางมิติที่ไม่ส่งผลต่อการพยากรณ์ภาพรวมระดับประเทศ จึงได้กำหนดแนวทางในการคัดกรองข้อมูลรบกวน (Noise) ออก เพื่อให้เหลือเฉพาะข้อมูลที่มีคุณภาพ

4.1.3 ผลการเตรียมข้อมูล (Data Preparation Results)

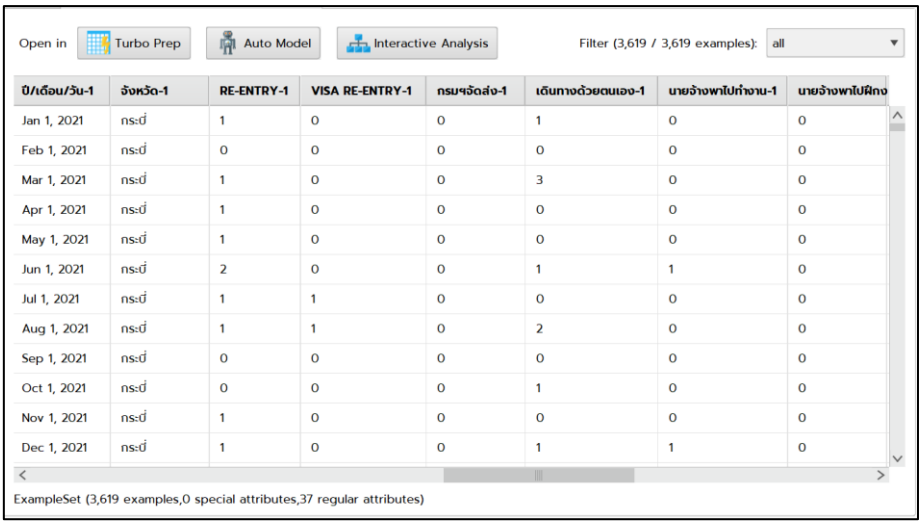
จากการนำข้อมูลเข้าสู่กระบวนการเตรียมข้อมูล (Data Preparation) ผ่านโปรแกรม Microsoft Excel , RapidMiner ทำให้ได้ชุดข้อมูลที่พร้อมสำหรับการวิเคราะห์ ดังนี้

4.1.3.1 ผลการแปลงข้อมูล (Aggregation) แปลงข้อมูลรายบุคคลกว่า 3 แสนระเบียน ให้เป็นข้อมูลอนุกรมเวลา (Time Series) โดยทำการรวบรวมข้อมูลเป็น รายเดือน จำแนกตามรายจังหวัด เพื่อให้โมเดลสามารถเรียนรู้รูปแบบการเดินทางของแรงงานในแต่ละพื้นที่ที่มีความแตกต่างกัน ซึ่งส่งผลให้ได้ชุดข้อมูลที่พร้อมสำหรับการวิเคราะห์รวมทั้งสิ้น 3,619 ระเบียน (จากช่วงเวลา 48 เดือน ในทุกจังหวัด) ทำให้เห็นรูปแบบแนวโน้มการขึ้นลงของแรงงานที่ชัดเจนขึ้น

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ID	เดือน	ปีปฏิทิน/วัน	จังหวัด	RE-ENTRY	VISA RE-ENTRY	กรมจัดส่ง	เดินทางด้วยตนเอง	นายจ้างพาไปทำงาน	นายจ้างพาไปฝึกงาน	บริษัทส่งตรงประจำเรือ	บริษัทจัดส่ง	ชาย	หญิง	15-25	26-35	36-45	46-60	>60	total
2	2021	1	2021-01-01	กระบี่	1	0	0	1	0	0	0	0	1	1	0	2	0	0	0	2
3	2021	2	2021-02-01	กระบี่	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2021	3	2021-03-01	กระบี่	1	0	0	3	0	0	0	0	2	2	0	3	1	0	0	4
5	2021	4	2021-04-01	กระบี่	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1
6	2021	5	2021-05-01	กระบี่	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1
7	2021	6	2021-06-01	กระบี่	2	0	0	1	1	0	0	0	2	2	0	1	3	0	0	4
8	2021	7	2021-07-01	กระบี่	1	1	0	0	0	0	0	0	2	0	0	1	1	0	0	2
9	2021	8	2021-08-01	กระบี่	1	1	0	2	0	0	0	0	2	2	0	0	3	1	0	4
10	2021	9	2021-09-01	กระบี่	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	2021	10	2021-10-01	กระบี่	0	0	0	1	0	0	0	1	1	1	2	0	0	0	0	2
12	2021	11	2021-11-01	กระบี่	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1
13	2021	12	2021-12-01	กระบี่	1	0	0	1	1	0	0	2	3	2	1	3	1	0	0	5
14	2021	1	2021-01-01	กรุงเทพมหานคร	18	0	0	8	8	5	0	15	34	20	9	16	16	10	3	54
15	2021	2	2021-02-01	กรุงเทพมหานคร	20	0	0	7	5	0	0	7	25	14	2	4	16	15	2	39
16	2021	3	2021-03-01	กรุงเทพมหานคร	19	0	2	5	16	1	0	12	37	18	1	20	21	12	1	55
17	2021	4	2021-04-01	กรุงเทพมหานคร	12	0	3	3	13	0	0	7	25	13	3	9	13	12	1	38
18	2021	5	2021-05-01	กรุงเทพมหานคร	21	2	1	10	11	0	0	8	35	18	4	18	13	18	0	53
19	2021	6	2021-06-01	กรุงเทพมหานคร	24	46	7	21	23	0	0	0	55	66	8	35	46	29	3	121
20	2021	7	2021-07-01	กรุงเทพมหานคร	7	38	4	20	41	1	0	1	62	50	8	39	26	38	1	112
21	2021	8	2021-08-01	กรุงเทพมหานคร	14	25	0	16	9	0	0	1	32	33	3	15	26	18	3	65
22	2021	9	2021-09-01	กรุงเทพมหานคร	3	24	1	12	5	0	0	5	25	25	1	17	17	14	1	50
23	2021	10	2021-10-01	กรุงเทพมหานคร	2	15	6	17	10	0	0	3	22	31	2	22	14	13	2	53
24	2021	11	2021-11-01	กรุงเทพมหานคร	6	15	2	22	15	0	0	6	34	32	4	16	26	17	3	66
25	2021	12	2021-12-01	กรุงเทพมหานคร	25	18	0	13	14	0	0	11	55	26	2	25	23	27	4	81
26	2021	1	2021-01-01	กาญจนบุรี	2	0	0	1	4	1	0	6	13	1	2	6	4	2	0	14
27	2021	2	2021-02-01	กาญจนบุรี	1	0	0	2	0	0	0	6	7	2	0	5	4	0	0	9
28	2021	3	2021-03-01	กาญจนบุรี	1	0	0	2	2	0	0	1	4	2	0	0	3	3	0	6
29	2021	4	2021-04-01	กาญจนบุรี	7	0	1	0	4	0	0	0	10	2	0	7	2	3	0	12

ภาพที่ 4.1 แสดงข้อมูลสถิติเดินทาง ปี 2564–2567

4.1.3.2 ผลการสร้างตัวแปรใหม่ (Feature Engineering) ผู้จัดทำได้สร้างตัวแปรเพิ่มเติมเพื่อให้โมเดลมีข้อมูลครบถ้วนสำหรับเรียนรู้แนวโน้ม โดยแบ่งเป็น 2 ส่วน คือ การสร้างตัวแปรย้อนหลัง (Lag Features) เนื่องจากเป็นข้อมูลอนุกรมเวลา (Time Series) ผู้จัดทำจึงสร้างตัวแปรย้อนหลัง 1 เดือน (Lag-1) ให้กับ ทุกปัจจัย ที่มี (ได้แก่ ยอดรวม, เพศ, ช่วงอายุ, และวิธีการเดินทาง) เพื่อให้โมเดลได้รู้สถานการณ์ในอดีตของทุกกลุ่ม สร้างตัวแปรระบุเวลา (Date Features) ได้แปลงข้อมูลวันที่ให้เป็น ดัชนีเดือน (Month_Key) เพื่อให้โมเดลสามารถจดจำรูปแบบของ "ฤดูกาล" (Seasonality) ได้ว่าในแต่ละเดือนมียอดการเดินทางสูงหรือต่ำ

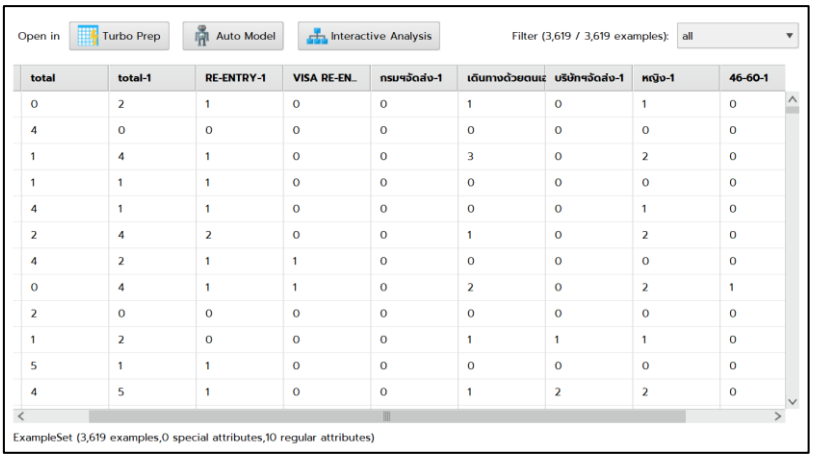


ปี/เดือน/วัน-1	จังหวัด-1	RE-ENTRY-1	VISA RE-ENTRY-1	กรมฯจัดส่ง-1	เดินทางด้วยตนเอง-1	นายจ้างพาไปทำงาน-1	นายจ้างพาไปฝึกง
Jan 1, 2021	กระบี่	1	0	0	1	0	0
Feb 1, 2021	กระบี่	0	0	0	0	0	0
Mar 1, 2021	กระบี่	1	0	0	3	0	0
Apr 1, 2021	กระบี่	1	0	0	0	0	0
May 1, 2021	กระบี่	1	0	0	0	0	0
Jun 1, 2021	กระบี่	2	0	0	1	1	0
Jul 1, 2021	กระบี่	1	1	0	0	0	0
Aug 1, 2021	กระบี่	1	1	0	2	0	0
Sep 1, 2021	กระบี่	0	0	0	0	0	0
Oct 1, 2021	กระบี่	0	0	0	1	0	0
Nov 1, 2021	กระบี่	1	0	0	0	0	0
Dec 1, 2021	กระบี่	1	0	0	1	1	0

ภาพที่ 4.2 แสดงข้อมูลผลการสร้างตัวแปรใหม่ (Feature Engineering)

4.1.3.3 ผลการคัดเลือกคุณลักษณะ (Feature Selection) จากการทดลองนำตัวแปรทั้งหมดเข้าสู่กระบวนการสร้างแบบจำลอง พบปัญหาเรื่องความซ้ำซ้อนของข้อมูล (Multicollinearity) โดยเฉพาะข้อมูลจำแนกตามเพศและช่วงอายุที่เมื่อรวมกันแล้วจะเท่ากับยอดรวม (Total) ซึ่งส่งผลให้แบบจำลองเกิดความสับสนในการคำนวณน้ำหนักความสำคัญ

ผู้จัดทำจึงได้ดำเนินการคัดเลือกคุณลักษณะ (Feature Selection) โดยพิจารณาจากค่าความสัมพันธ์ (Correlation) และนัยสำคัญทางสถิติ ผลการวิเคราะห์พบว่า ตัวแปรส่วนใหญ่มีความซ้ำซ้อนและไม่มีนัยสำคัญต่อการพยากรณ์ จึงได้ทำการตัดออก อย่างไรก็ตาม พบว่ามีตัวแปรบางกลุ่มที่มีความสัมพันธ์กับยอดรวมอย่างมีนัยสำคัญและช่วยเพิ่มความแม่นยำให้กับแบบจำลองได้ ได้แก่ ยอดรวมย้อนหลัง 1 เดือน (Total–1), ข้อมูลเพศหญิงย้อนหลัง (หญิง–1), และข้อมูลช่วงอายุ 46–60 ปี (46–60–1) รวมถึงตัวแปรด้านวิธีการเดินทาง (กรมฯ/บริษัทฯ/เดินทางด้วยตนเอง) ดังนั้น ผู้จัดทำจึงคัดเลือกเฉพาะตัวแปรที่มีนัยสำคัญเหล่านี้ไว้ และตัดตัวแปรอื่น ๆ ที่ซ้ำซ้อนออก (เช่น เพศชาย หรือช่วงอายุอื่น ๆ) ซึ่งส่งผลให้แบบจำลองมีความกระชับและลดค่าความคลาดเคลื่อนได้ดียิ่งขึ้น


Open in Turbo Prep Auto Model Interactive Analysis Filter (3,619 / 3,619 examples): all

total	total-1	RE-ENTRY-1	VISA RE-EN..	กรมฯจัดส่ง-1	เดินทางด้วยตนเอ	บริษัทฯจัดส่ง-1	หญิง-1	46-60-1
0	2	1	0	0	1	0	1	0
4	0	0	0	0	0	0	0	0
1	4	1	0	0	3	0	2	0
1	1	1	0	0	0	0	0	0
4	1	1	0	0	0	0	1	0
2	4	2	0	0	1	0	2	0
4	2	1	1	0	0	0	0	0
0	4	1	1	0	2	0	2	1
2	0	0	0	0	0	0	0	0
1	2	0	0	0	1	1	1	0
5	1	1	0	0	0	0	0	0
4	5	1	0	0	1	2	2	0

ExampleSet (3,619 examples,0 special attributes,10 regular attributes)

ภาพที่ 4.3 แสดงข้อมูลผลการคัดเลือกคุณลักษณะ (Feature Selection)

4.1.4 ผลการสร้างแบบจำลอง (Modeling)

จากการดำเนินการเตรียมและทำความสะอาดข้อมูล (Data Cleaning) ซึ่งมีข้อมูลดิบเริ่มต้นจำนวนกว่า 324,497 แถวซึ่งเป็นชนิดรายบุคคล จึงได้ทำการคัดกรองจัดการค่าว่างและแปลงข้อมูลจนได้ชุดข้อมูลที่สมบูรณ์พร้อมสำหรับการวิเคราะห์คือ จำนวน 3,619 แถว เพื่อให้การประเมินประสิทธิภาพของแบบจำลองมีความน่าเชื่อถือและป้องกันภาวะ Overfitting จึงได้ดำเนินการแบ่งชุดข้อมูล (Data Splitting) ออกเป็น 2 ส่วน ด้วยสัดส่วน 70:30 โดยมีรายละเอียดจำนวนแถวข้อมูลดังนี้

4.1.4.1 ข้อมูลสำหรับเรียนรู้ (Training Set) จำนวนร้อยละ 70 สำหรับใช้ในการสร้างและสอนโมเดล มีจำนวนทั้งสิ้น 2,533 แถว

4.1.4.2 ข้อมูลสำหรับทดสอบ (Testing Set) จำนวนร้อยละ 30 สำหรับใช้ในการวัดผลและประเมินความแม่นยำ มีจำนวนทั้งสิ้น 1,086 แถว

โดยชุดข้อมูลดังกล่าวได้ถูกนำไปใช้ในการสร้างแบบจำลองการพยากรณ์แรงงานไทย จำนวน 4 โมเดลหลัก ได้แก่

1) Multiple Linear Regression (MLR) แบบจำลอง MLR สามารถสะท้อนแนวโน้มการเปลี่ยนแปลงของจำนวนแรงงานได้ชัดเจน ทั้งในช่วงที่เพิ่มขึ้นและลดลง เนื่องจากข้อมูลมีลักษณะเป็นแนวโน้มตามเวลา ทำให้แบบจำลองเชิงเส้นสามารถเรียนรู้รูปแบบได้ดี จากการทดลองพบว่า MLR ให้ผลการพยากรณ์ใกล้เคียงกับค่าจริงมากที่สุดเมื่อเทียบกับโมเดลอื่น จึงถือว่าเป็นแบบจำลองที่เหมาะสมที่สุดสำหรับข้อมูลชุดนี้

2) Gradient Boosted Trees (GBT) แบบจำลอง GBT สามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนและไม่เป็นเชิงเส้นได้ดี โดยค่าพยากรณ์มีความใกล้เคียงกับข้อมูลจริงในหลาย

ช่วงเวลา แต่ในบางช่วงที่ข้อมูลมีแนวโน้มชัดเจน แบบจำลองยังมีความผันผวนมากกว่า MLR เล็กน้อย

3) Random Forest Regression ผลลัพธ์ค่อนข้างเสถียร สามารถลดผลกระทบจากค่าผิดปกติได้ดี ค่าพยากรณ์มีความใกล้เคียงกับค่าจริงในระดับหนึ่ง แต่ยังไม่แม่นยำเท่ากับ MLR ในภาพรวม

4) k–Nearest Neighbors (k–NN) พยากรณ์ได้ดีในช่วงเวลาที่มีรูปแบบใกล้เคียงกับข้อมูลในอดีต แต่ผลลัพธ์มีความผันผวนขึ้นอยู่กับค่า k ที่กำหนดซึ่งได้ทำการตั้งค่าผ่าน Parameter ที่ดีที่สุดแล้ว แต่ผลลัพธ์ทำให้ความแม่นยำโดยรวมยังด้อยกว่าแบบจำลองอื่น สรุปจากการสร้างแบบจำลองทั้ง 4 เทคนิค พบว่าแบบจำลอง Multiple Linear Regression (MLR) ให้ผลการพยากรณ์ดีที่สุดและมีความเหมาะสมกับลักษณะข้อมูลแรงงานไทยในต่างประเทศมากที่สุด โดยรายละเอียดเชิงตัวเลขจะนำเสนอดังตาราง 4.1.5 การประเมินผลต่อไป

4.1.5 ผลการประเมินผล (Evaluation)

ผลการเปรียบเทียบประสิทธิภาพของโมเดลพยากรณ์ทั้ง 4 โมเดล ได้แก่ 1) Multiple Linear Regression 2) Gradient Boosted Trees (GBT) 3) Random Forest Regression และ 4) k–Nearest Neighbors (kNN) โดยโมเดลเหล่านี้จะถูกวัดประสิทธิภาพโดยใช้ค่าความผิดพลาด RMSE (Root Mean Square Error) และ MAE (Mean Absolute Error) ซึ่งเป็นตัวชี้วัดมาตรฐานในการประเมินความแม่นยำของแบบจำลอง โดยพิจารณาจากผลต่างระหว่างค่าพยากรณ์และค่าจริง เพื่อค้นหาโมเดลที่แม่นยำที่สุด ผู้วิจัยได้นำค่าความคลาดเคลื่อนบน ชุดข้อมูลทดสอบ (Testing Set) มาเปรียบเทียบกัน ดังตารางที่ 4.1

ตารางที่ 4.1 ตารางเปรียบเทียบประสิทธิภาพระหว่างโมเดล

ตัวแบบ	RMSE	MAE
Multiple Linear Regression	67.683	29.764
Gradient Boosted Trees (GBT)	87.871	33.447
Random Forest Regression	81.886	32.465
k–Nearest Neighbors (kNN)	100.867	36.957

จากตารางที่ 4.1 จากการเปรียบเทียบประสิทธิภาพ พบว่าแบบจำลอง Multiple Linear Regression (MLR) ให้ค่าความคลาดเคลื่อน (RMSE และ MAE) ต่ำที่สุด สามารถนำไป

ประยุกต์ใช้กับข้อมูลใหม่ (Generalization) ได้ดีกว่าโมเดลอื่น ๆ ที่พบค่าความคลาดเคลื่อนสูงกว่าในชุดข้อมูลทดสอบ ผู้ศึกษาจึงเลือกใช้ MLR เป็นต้นแบบหลักในการพยากรณ์

4.1.6 ผลการนำไปใช้งาน (Deployment) จากการดำเนินการในขั้นตอนการประเมินผล (Evaluation) ได้คัดเลือกแบบจำลอง Multiple Linear Regression (MLR) เป็นแบบจำลองที่เหมาะสมที่สุด เนื่องจากมีค่าความคลาดเคลื่อนต่ำที่สุด (RMSE = 67.683 และ MAE = 29.764) และมีความเสถียรในการทดสอบกับชุดข้อมูลใหม่

ในขั้นตอนการนำไปใช้งาน (Deployment) ผู้ศึกษาได้นำผลลัพธ์ที่ได้จากการวิเคราะห์ผ่านโปรแกรม RapidMiner มาเป็นแนวทางการใช้งานจริง ดังนี้

4.1.6.1 รูปแบบพยากรณ์และประยุกต์ใช้งาน การสร้างโมเดลการพยากรณ์โดยใช้การถดถอยเชิงเส้น Multiple Linear Regression (MLR) ของตัวแบบพยากรณ์ เพื่อให้สามารถนำผลไปประยุกต์ใช้ในการพยากรณ์แรงงานไทยส่งออกไปต่างประเทศในอนาคตได้ จึงได้แปลงค่าสัมประสิทธิ์ (Coefficients) จากแบบจำลอง MLR ให้อยู่ในรูปของสมการทางคณิตศาสตร์ ดังนี้

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value
RE-ENTRY-1	-0.281	0.167	-0.033	0.475	-1.681	0.093
VISA RE-ENTRY-1	0.326	0.145	0.056	0.302	2.255	0.024
กรมฯจัดส่ง-1	1.444	0.180	0.162	0.451	8.010	0.000
เดินทางด้วยตนเอง-1	-1.577	0.100	-0.410	0.274	-15.827	0
บริษัทฯจัดส่ง-1	0.809	0.097	0.198	0.325	8.340	0.000
หญิง-1	1.939	0.182	0.342	0.179	10.678	0
46-60-1	1.150	0.091	0.378	0.204	12.593	0
(Intercept)	7.697	2.632	?	?	2.924	0.003

ภาพที่ 4.4 แสดงค่าสัมประสิทธิ์ (Coefficients)

จากภาพที่ 4.4 แสดงค่าสัมประสิทธิ์ของตัวแปรต่าง ๆ ในสมการ Linear Regression โดยสมการนี้ใช้ในการพยากรณ์ตัวแปรตาม (ในกรณีนี้คือ total หรือจำนวนแรงงานรวม) จากตัวแปรอิสระหลายตัว เช่น เพศ, วิธีการจัดส่ง, ช่วงอายุ และประเภทวีซ่า โดยรูปแบบของสมการ มีดังนี้

$$\hat{Y} = 7.697 - 0.281(\text{RE} - \text{ENTRY}) + 0.326(\text{VISA RE} - \text{ENTRY}) + 1.444(\text{กรมจัดส่ง}) - 1.577(\text{เดินทางด้วยตนเอง}) + 0.809(\text{บริษัทจัดส่ง}) + 1.939(\text{หญิง}) + 1.150(\text{อายุ } 46-60 \text{ ปี})$$

โดยที่

$\hat{Y}$ คือ ค่าพยากรณ์จำนวนแรงงานรวม

ตัวแปรอิสระเป็นค่าปริมาณจริง จากจำนวนแรงงานในแต่ละกลุ่ม

ค่าสัมประสิทธิ์สะท้อนการเปลี่ยนแปลงของจำนวนแรงงานรวมเมื่อค่าตัวแปรอิสระเพิ่มขึ้น 1 หน่วย ภายใต้เงื่อนไขที่ตัวแปรอื่นคงที่

การแปลความหมายจากสมการ มีรายละเอียด ดังนี้

จากสมการพยากรณ์สามารถตีความค่าสัมประสิทธิ์ของตัวแปรแต่ละตัวได้ดังนี้ โดยกำหนดให้ตัวแปรอื่นคงที่

ค่า Intercept เท่ากับ 7.697 หมายความว่า หากค่าตัวแปรอิสระทั้งหมดมีค่าเป็นศูนย์ จะได้ค่าพยากรณ์จำนวนแรงงานรวมเท่ากับ 7.697 หน่วย

ตัวแปร RE-ENTRY มีค่าสัมประสิทธิ์เท่ากับ -0.281 หมายความว่า หากจำนวนแรงงานที่เป็น RE-ENTRY เพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมลดลงเฉลี่ย 0.281 หน่วย

ตัวแปร VISA RE-ENTRY มีค่าสัมประสิทธิ์เท่ากับ 0.326 หมายความว่า หากจำนวนแรงงานที่เป็น VISA RE-ENTRY เพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมเพิ่มขึ้นเฉลี่ย 0.326 หน่วย

ตัวแปร กรมจัดส่ง มีค่าสัมประสิทธิ์เท่ากับ 1.444 หมายความว่า หากจำนวนแรงงานที่จัดส่งโดยกรมจัดหางานเพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมเพิ่มขึ้นเฉลี่ย 1.444 หน่วย

ตัวแปร เดินทางด้วยตนเอง มีค่าสัมประสิทธิ์เท่ากับ -1.577 หมายความว่า หากจำนวนแรงงานที่เดินทางด้วยตนเองเพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมลดลงเฉลี่ย 1.577 หน่วย

ตัวแปร บริษัทจัดส่ง มีค่าสัมประสิทธิ์เท่ากับ 0.809 หมายความว่า หากจำนวนแรงงานที่จัดส่งโดยบริษัทเอกชนเพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมเพิ่มขึ้นเฉลี่ย 0.809 หน่วย

ตัวแปร หญิง มีค่าสัมประสิทธิ์เท่ากับ 1.939 หมายความว่า หากจำนวนแรงงานเพศหญิงเพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมเพิ่มขึ้นเฉลี่ย 1.939 หน่วย

ตัวแปร อายุ 46–60 ปี มีค่าสัมประสิทธิ์เท่ากับ 1.150 หมายความว่า หากจำนวนแรงงานในช่วงอายุ 46–60 ปีเพิ่มขึ้น 1 หน่วย จะทำให้จำนวนแรงงานรวมเพิ่มขึ้นเฉลี่ย 1.150 หน่วย

การพยากรณ์จะดำเนินการโดยการแทนค่าของแต่ละตัวแปรตามข้อมูลที่มี จากนั้นนำค่าที่ได้คูณกับค่าสัมประสิทธิ์ (Coefficient) แล้วนำมาบวกหรือลบเข้ากับค่าคงที่ (Intercept = 7.697) เพื่อหาค่าพยากรณ์สุดท้าย

ผลการวัดประสิทธิภาพของโมเดล (Performance Vector) จากการทดสอบแบบจำลอง พบว่ามีค่า Root Mean Squared Error (RMSE) หรือค่าความผิดพลาดรากที่สองของค่าเฉลี่ยกำลังสอง เท่ากับ 67.683 และค่า Mean Absolute Error (MAE) หรือค่าความผิดพลาดเชิงสัมบูรณ์เฉลี่ย เท่ากับ 29.764 ซึ่งถือว่าเป็นโมเดลที่มีประสิทธิภาพสูงที่สุดเมื่อเทียบกับแบบจำลองอื่น ๆ เนื่องจากมีความคลาดเคลื่อนต่ำ แสดงถึงการคาดการณ์ที่แม่นยำและสอดคล้องกับข้อมูลจริง ดังแสดงในภาพที่ 4.5

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 67.683 +/- 0.000
absolute_error: 29.764 +/- 60.787
```

ภาพที่ 4.5 แสดงผลลัพธ์ Performance Vector ของการถดถอยเชิงเส้น (Multiple Linear Regression)

4.1.6.2 ผลลัพธ์จากการพยากรณ์ด้วยตัวแบบจะถูกนำมาแสดงในรูปแบบกราฟต่าง ๆ เพื่อแสดงผลผ่านบนเว็บไซต์ คือ นำผลลัพธ์จาก Multiple Linear Regression Model ที่ใช้พยากรณ์ค่าตัวแปร total บนชุดข้อมูลที่ประกอบด้วยตัวแปรเดือน แสดงดังภาพที่ 4.6

Open in Turbo Prep | Auto Model | Interactive Analysis — Filter (1,086 / 1,086 examples): all

Row No.	ปี/เดือน/วัน	total	prediction(t..	RE-ENTRY-1	VISA RE-EN..	กรมฯจัดส่ง-1	เดินทางด้วยตนเอ	บริษัทฯจัดส่ง-
1	8	6	14.000	0	1	2	0	0
2	9	5	14.213	1	1	1	0	0
3	10	4	10.163	1	0	0	0	1
4	11	3	8.565	1	0	0	0	0
5	12	4	10.287	0	2	0	0	0
6	2	6	11.032	2	0	0	0	1
7	3	9	12.214	3	0	1	0	2
8	4	12	13.505	5	0	1	2	1
9	5	4	13.605	6	0	1	3	2
10	6	13	10.932	0	0	0	0	4

ExampleSet (1,086 examples,2 special attributes,8 regular attributes)

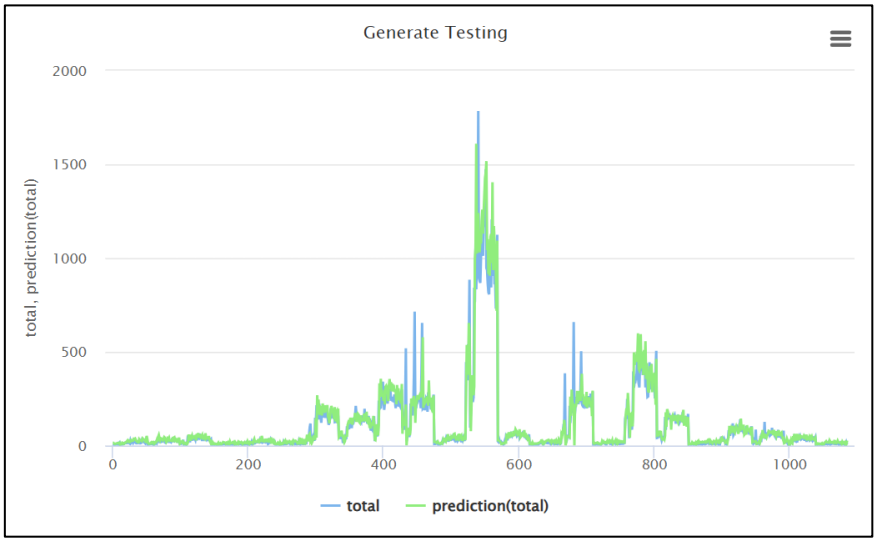
ภาพที่ 4.6 แสดงผลลัพธ์จาก Multiple Linear Regression Model

นำข้อมูลที่ได้ไปเข้า Microsoft Excel เพื่อสะดวกในการใช้งานของข้อมูล แสดงดังภาพที่ 4.7

	A	B
1	**total**	**predict**
2	6	13.9997
3	5	14.21295
4	4	10.16274
5	3	8.565371
6	4	10.28721
7	6	11.03155
8	9	12.2145
9	12	13.5048
10	4	13.6053
11	13	10.93196
12	17	14.73441
13	4	0.789053
14	9	8.745769
15	7	9.460842
16	5	15.28777
17	9	13.5407
18	11	9.198789
19	10	12.66747
20	15	15.60385
21	14	15.16022

ภาพที่ 4.7 แสดงค่า total และค่า predict ของ Multiple Linear Regression

เพื่อให้เห็นภาพประสิทธิภาพของแบบจำลอง Multiple Linear Regression ในการใช้งานจริง จึงได้นำข้อมูลผลลัพธ์การพยากรณ์ (Predicted Values) มาเปรียบเทียบกับข้อมูลจำนวนแรงงานจริง (Actual Values) ในช่วงเวลาทดสอบ โดยแสดงผลในรูปแบบกราฟเส้น ดังภาพที่ 4.8 โดยนำชุดข้อมูล Linear Regression ไปวิเคราะห์และสร้าง Visualization เพื่อเปรียบเทียบโมเดลพยากรณ์ แสดงดังภาพที่ 4.8



ภาพที่ 4.8 แสดงผลการพยากรณ์ถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression)

ผลการวิเคราะห์เปรียบเทียบ จากภาพที่ 4.8 แสดงความสัมพันธ์ระหว่างเส้นกราฟข้อมูลจริง เส้นสีฟ้า และเส้นกราฟค่าพยากรณ์ เส้นสีเขียว โดยวิเคราะห์ออกมาได้ ดังนี้ คือ แบบจำลองสามารถพยากรณ์ทิศทางของข้อมูลได้อย่างแม่นยำ โดยเส้นค่าพยากรณ์มีการเคลื่อนที่สอดคล้องกับข้อมูลจริงตลอดช่วงเวลาทดสอบ คือ มีแนวโน้มเพิ่มขึ้นในช่วงครึ่งปีแรก และปรับตัวลดลงในช่วงปลายปี แสดงให้เห็นว่าแบบจำลองสามารถจับฤดูกาล (Seasonality) หรือวัฏจักรของการส่งออกแรงงานได้

4.2 ผลการวิเคราะห์และออกแบบระบบ

4.2.1 ผลการวิเคราะห์และออกแบบระบบฐานข้อมูล

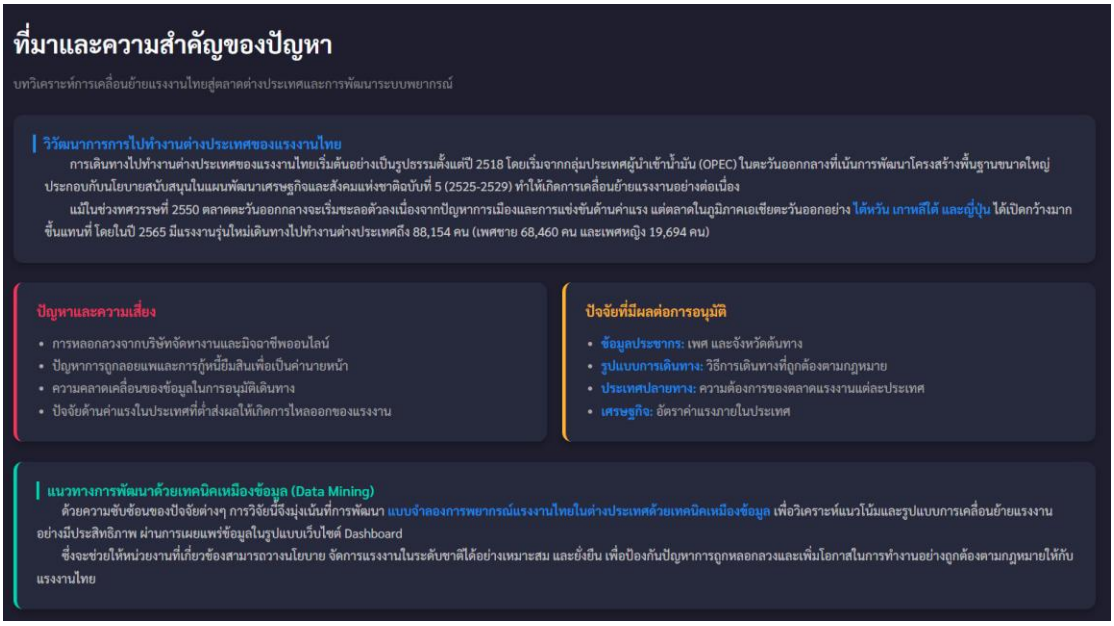
ระบบการจัดการฐานข้อมูลแรงงานไทยที่เดินทางไปทำงานที่ต่างประเทศระบบได้มีการแสดงข้อมูลการเดินทางไปทำงานในต่างประเทศของแรงงานไทยมีระบบฟังก์ชั่นหลักได้แก่ แสดงจำนวนข้อมูลแรงงานรวมทั้งหมด จังหวัดยอดนิยมสูงสุด วิธีการเดินทางหลัก เลือกปีข้อมูล

4.2.2 ด้านผู้ใช้งานทั่วไป

1) หน้าแสดงสถิติแรงงานไทยในต่างประเทศ

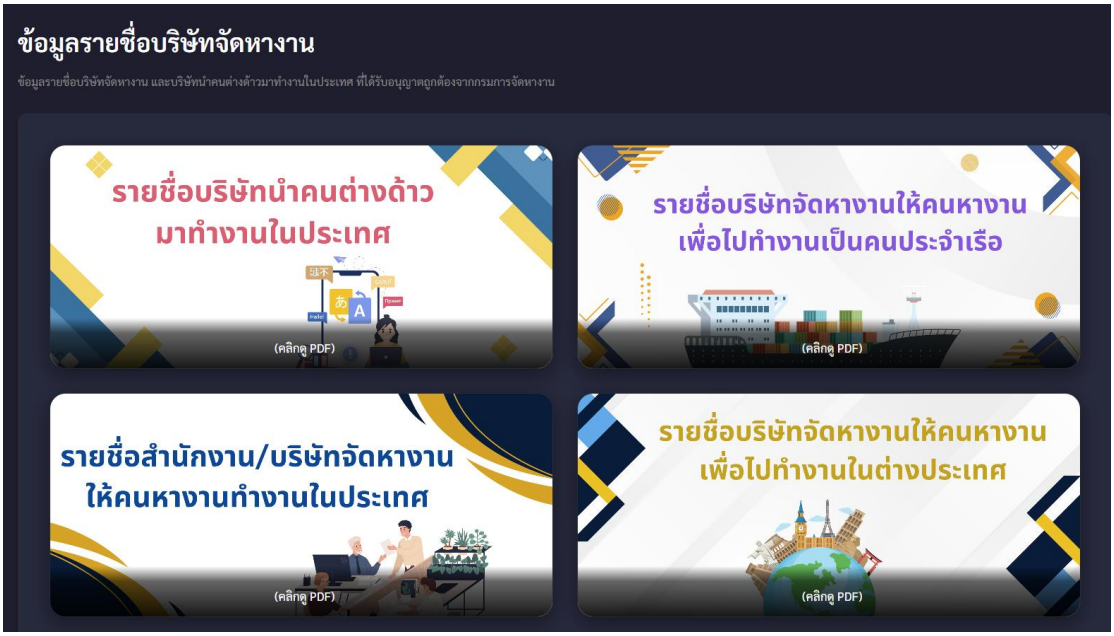
ภาพที่ 4.9 หน้าแสดงสถิติแรงงานไทยในต่างประเทศโดยนำข้อมูลจริงมาวิเคราะห์และสรุปผลในรูปแบบกราฟและตัวชี้วัด เพื่อให้เห็นภาพรวมของสถานการณ์แรงงานได้อย่างเป็นระบบ

2) หน้าแสดงที่มาและความสำคัญ



ภาพที่ 4.10 หน้าที่มาและความสำคัญของปัญหาของระบบ โดยอธิบายภาพรวมของแรงงานไทยที่ไปทำงานต่างประเทศ

3) หน้าแสดงข้อมูลรายชื่อบริษัทจัดหางาน



ภาพที่ 4.11 หน้าแสดงข้อมูลรายชื่อบริษัทจัดหางานที่ได้รับอนุญาตอย่างถูกต้อง

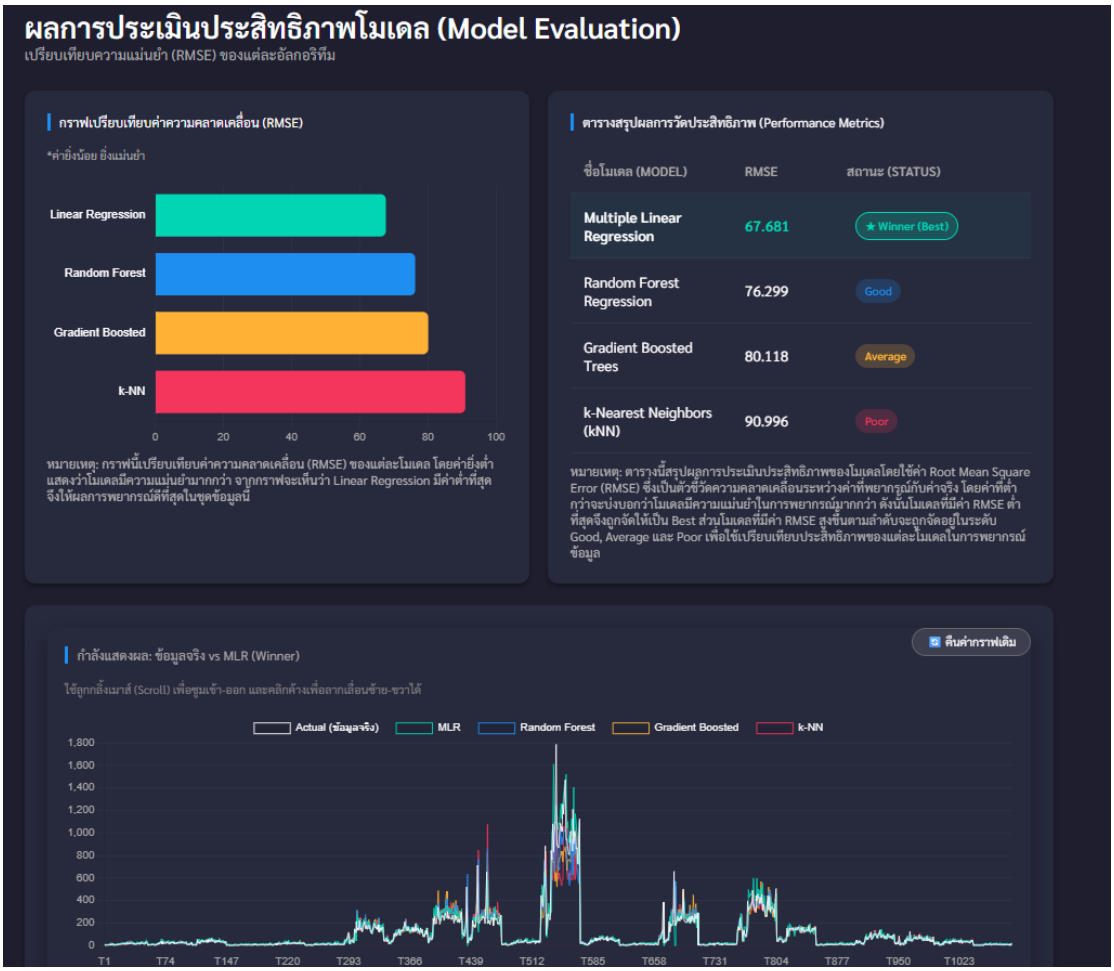
4) หน้าแสดงอัตราค่าจ้าง

ภาพที่ 4.12 หน้าแสดงข้อมูลอัตราค่าจ้างขั้นต่ำของประเทศไทย ปี 2568 โดยอ้างอิงจากประกาศของคณะกรรมการค่าจ้าง ซึ่งมีผลบังคับใช้ตั้งแต่วันที่ 1 กรกฎาคม 2568 เป็นต้นไป

5) หน้าแสดงผลของระบบพยากรณ์แรงงานไทยไปต่างประเทศ

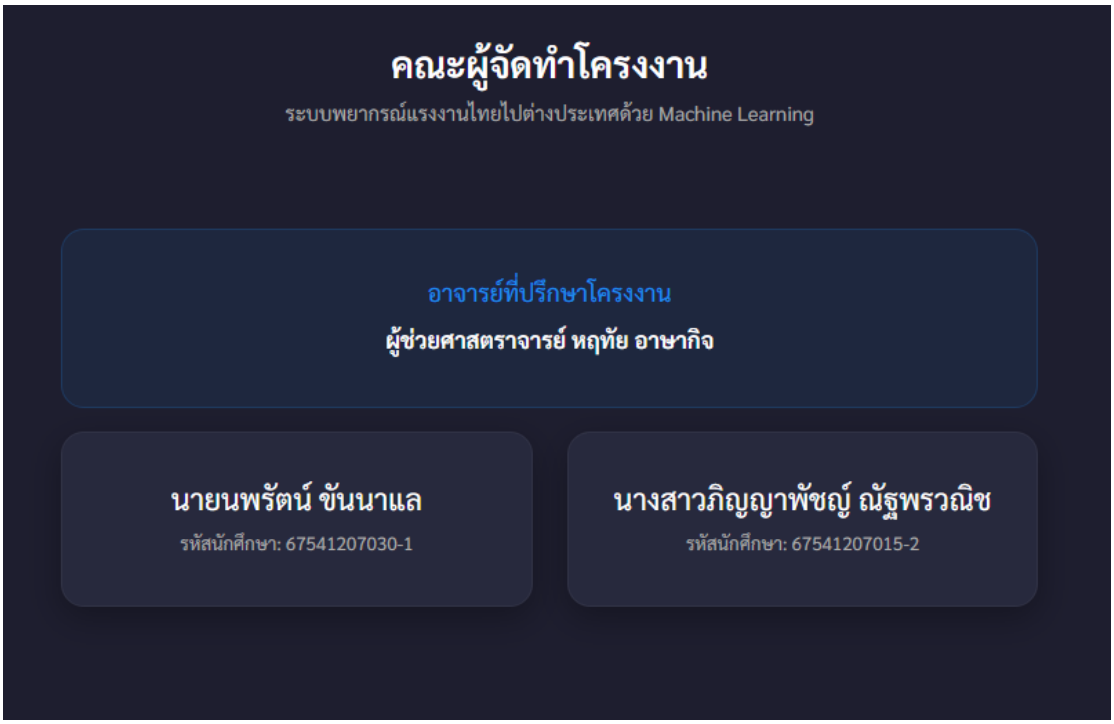
ภาพที่ 4.13 หน้าแสดงผลของระบบพยากรณ์แรงงานไทยไปต่างประเทศ โดยเน้นการนำเสนอผลการวิเคราะห์ข้อมูลและผลลัพธ์จากแบบจำลองเชิงสถิติ

6) หน้าแสดงผลการประเมินประสิทธิภาพของโมเดล



ภาพที่ 4.14 หน้าแสดงผลการประเมินประสิทธิภาพของโมเดล (Model Evaluation) โดยมีวัตถุประสงค์เพื่อเปรียบเทียบความแม่นยำของอัลกอริทึมแต่ละแบบ และเลือกโมเดลที่เหมาะสมที่สุดสำหรับการพยากรณ์

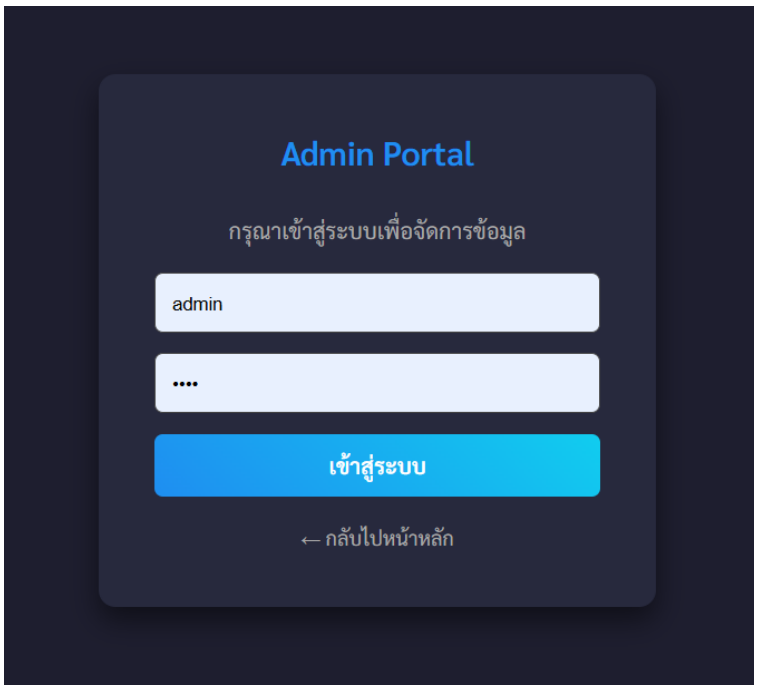
7) หน้าแสดงข้อมูลผู้จัดทำและอาจารย์ที่ปรึกษา



ภาพที่ 4.15 หน้าแสดงข้อมูลผู้จัดทำและอาจารย์ที่ปรึกษา

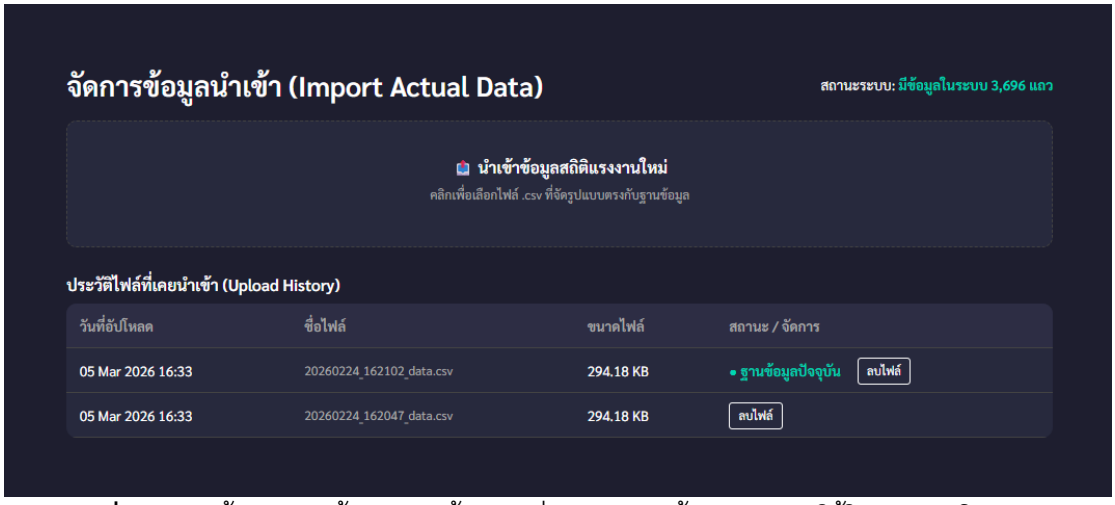
4.2.2 ด้านผู้ดูแลระบบ

8. หน้าเข้าสู่ระบบจัดการ สำหรับผู้ดูแลระบบ



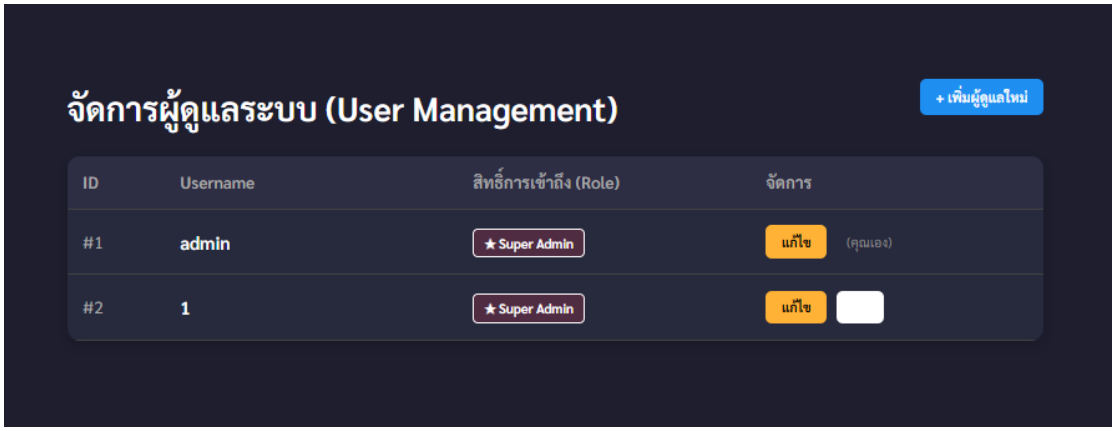
ภาพที่ 4.16 หน้าแสดงการเข้าสู่ระบบ

9. หน้าการจัดการข้อมูล



ภาพที่ 4.17 หน้าจัดการข้อมูลนำเข้า เป็นส่วนสำหรับผู้ดูแลระบบ ใช้ในการอัปโหลดและจัดการข้อมูลแรงงานเข้าสู่ระบบ

10. หน้าการจัดการผู้ใช้



ภาพที่ 4.18 หน้าจัดการผู้ดูแลระบบ เป็นส่วนสำหรับกำหนดและบริหารจัดการบัญชีผู้ใช้งานในระบบ โดยเฉพาะผู้ที่มีสิทธิ์เข้าถึงข้อมูลและฟังก์ชันต่าง ๆ ภายในระบบ