

ในกระบวนการนี้ประกอบด้วย 6 ขั้นตอน

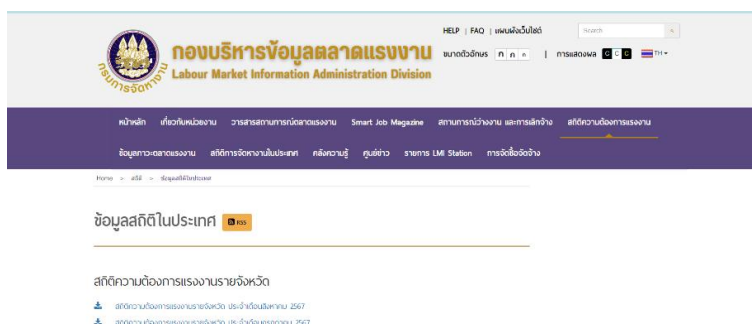
3.1.1 ความเข้าใจในธุรกิจ (Business Understanding)

ผู้วิเคราะห์มุ่งเน้นไปที่การทำความเข้าใจกระบวนการทางข้อมูลและระบบวิเคราะห์ข้อมูล จากสิ่งรอบตัวการสัมภาษณ์หรือรับฟังปัญหา รวมถึงการมองหาปัญหาจากเรื่องต่าง ๆ เพื่อที่จะนำผลการวิเคราะห์ข้อมูลไปใช้ให้เกิดประโยชน์ โดยจัดลำดับความสำคัญ และกำหนด วัตถุประสงค์ที่จะนำไปสู่รูปแบบการวิเคราะห์ข้อมูล ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูล ฐานข้อมูลขนาดใหญ่และวางแผนในการดำเนินการโดยการวิเคราะห์ข้อมูลทาง Data Mining หารูปแบบความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลของข้อมูลจำนวนผู้ว่างงาน และกำลังแรงงานรวม เพื่อที่จะนำผลการวิเคราะห์ข้อมูลไปใช้ ประโยชน์ความต้องการทั้งหมดจะนำมาจัดลำดับความสำคัญ และกำหนดวัตถุประสงค์ที่จะ นำไปสู่รูปแบบการวิเคราะห์ข้อมูล

3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)

ขั้นตอนการจัดเก็บและรวบรวมข้อมูล ผู้วิเคราะห์ข้อมูลต้องพิจารณาตรวจสอบ ความถูกต้องของข้อมูลที่ได้รับ โดยเลือกว่าจะใช้ข้อมูลทั้งหมดหรือบางส่วนในการวิเคราะห์ให้ สอดรับกับ วัตถุประสงค์ที่กำหนดไว้ วิจัยสำรวจภาพรวมทั้งหมดก่อนนำมาใช้ในการวิเคราะห์ข้อมูล จากเว็บไซต์ www.doe.go.th ซึ่งกองบริหารข้อมูลตลาดแรงงาน ให้บริการข้อมูล และข่าวสารแก่ประชาชนใน ลักษณะข้อมูลเปิดเผย (Open Data) ผู้วิเคราะห์ข้อมูลจึงได้ ดำเนินการตามขั้นตอนดังนี้

1) ผู้วิเคราะห์ข้อมูลรวบรวมข้อมูลจาก เว็บไซต์ www.doe.go.th แล้วคลิกแถบเมนู สถิติความต้องการแรงงาน สามารถคลิกเดือนที่ต้องการแล้วดาวน์โหลด เพื่อเริ่มทำการดาวน์โหลด ข้อมูลเก็บรวบรวมข้อมูล



ภาพที่ 3.2 แสดงขั้นตอนการหา Open Data จากเว็บไซต์

ที่มา: <https://www.doe.go.th/prd/lmia/statistic/>

2) เลือกข้อมูลที่จะใช้คือ ข้อมูลความต้องการแรงงานรายจังหวัด ตั้งแต่ปีพ.ศ.

2564 – 2566 จาก www.doe.go.th

- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบประมาณ 2566
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบประมาณ 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯพิเศษ 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯปกติ 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯพิเศษ 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯปกติ 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯพิเศษ 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2565
- ๓. สถิติความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯปกติ 2565
- ๓. สถานการณ์แรงงานภาคความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2564
- ๓. สถานการณ์แรงงานภาคความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2564
- ๓. สถานการณ์แรงงานภาคความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2564
- ๓. สถานการณ์แรงงานภาคความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯปกติ 2564
- ๓. สถานการณ์แรงงานภาคความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2564
- ๓. สถานการณ์แรงงานภาคความต้องการแรงงาน ชนิดหัวปี ประจำปีงบฯฉุกเฉิน 2564

ภาพที่ 3.3 แสดงขั้นตอนการหา Open Data จากเว็บไซต์

ที่มา: <https://www.doe.go.th/prd/lmia/statistic>

3) ตรวจสอบข้อมูลดาวน์โหลดโดยพิจารณาจากข้อมูลทั้งหมด มีจำนวนข้อมูลไฟล์ทั้งหมด 288 ไฟล์ ซึ่ง 1 ไฟล์ ประกอบไปด้วย 10 แอดทริบิวต์

จังหวัด	ประชากรอายุ 15 ปีขึ้นไป	กำลังแรงงานรวม					ผู้ไม่อยู่ในกำลังแรงงาน				
		รวม	ผู้มีงานทำ	ผู้ว่างงาน	ผู้ที่อยู่ภาค	รวม	ทำงานบ้าน	เรียนหนังสือ	เด็ก ขาด บวช พิการ	อื่น ๆ	
รวม	58,661,185	40,088,586	39,565,992	491,409	31,185	18,572,599	5,111,012	4,558,354	6,963,369	1,939,864	
1. กรุงเทพมหานคร	7,859,114	5,699,929	5,634,878	65,051	0	2,159,185	581,516	450,011	750,501	377,157	
2. สมุทรปราการ	2,075,236	1,425,096	1,409,201	15,895	0	650,140	187,804	121,789	251,556	88,991	
3. นครปฐม	1,643,154	1,172,113	1,157,263	14,850	0	471,041	106,546	123,707	167,153	73,635	
4. ปทุมธานี	1,646,354	1,091,811	1,082,728	9,083	0	554,543	189,982	128,168	168,602	67,791	
5. พระนครศรีอยุธยา	785,469	532,422	520,392	12,030	0	253,047	64,127	54,270	103,351	31,299	
6. อ่างทอง	208,690	128,029	126,366	1,663	0	80,661	24,731	15,615	27,464	12,851	
7. สทูลบุรี	658,692	452,591	441,871	10,720	0	206,101	54,653	53,149	72,808	25,491	
8. สิงห์บุรี	157,204	99,943	97,076	2,867	0	57,261	19,972	11,807	19,051	6,431	
9. ชัยนาท	245,329	168,758	164,332	4,426	0	76,571	22,763	15,606	32,469	5,733	
10. สระบุรี	652,803	442,680	431,705	10,975	0	210,123	54,867	45,743	86,655	22,858	
11. ชลบุรี	1,641,870	1,101,655	1,098,137	3,518	0	540,215	144,487	108,057	197,072	90,599	
12. ระยอง	900,120	686,312	675,556	10,756	0	213,808	75,110	54,093	69,166	15,439	
13. จันทบุรี	471,109	369,632	368,003	1,629	0	101,477	29,428	25,674	40,317	6,058	
14. ตราด	228,991	170,103	169,117	986	0	58,888	19,504	13,139	20,584	5,661	
15. ฉะเชิงเทรา	743,767	525,528	522,946	2,582	0	218,239	83,760	54,011	65,324	15,144	
16. ปราจีนบุรี	543,085	397,371	389,049	8,322	0	145,714	59,115	30,176	38,202	18,221	
17. นครนายก	218,550	149,417	148,994	344	79	69,133	20,961	14,162	23,038	10,972	
18. สระแก้ว	549,194	374,930	368,878	5,449	603	174,264	64,662	32,572	52,283	24,747	
19. ราชบุรี	679,976	472,734	469,260	3,474	0	207,242	55,171	43,221	88,913	19,937	
20. กาญจนบุรี	696,667	487,620	482,738	4,882	0	209,047	62,604	41,116	85,083	20,244	
21. สุพรรณบุรี	709,538	485,531	482,641	2,890	0	224,007	48,340	57,586	99,845	18,236	
22. นครปฐม	1,111,574	828,149	819,964	8,185	0	283,425	80,677	84,954	84,774	33,020	
23. นครสวรรค์	987,349	758,576	749,506	9,070	0	228,773	65,147	58,620	67,752	37,254	
24. นครราชสีมา	148,235	102,783	101,939	844	0	45,452	16,297	10,091	14,532	4,532	
25. เพชรบุรี	421,771	296,282	292,848	3,434	0	125,489	29,050	32,386	47,230	16,823	
26. ประจวบคีรีขันธ์	406,346	270,470	268,849	1,621	0	135,876	39,939	29,687	46,520	19,730	
27. เชียงใหม่	1,522,971	1,043,916	1,023,177	20,739	0	479,055	101,978	128,693	197,293	51,091	

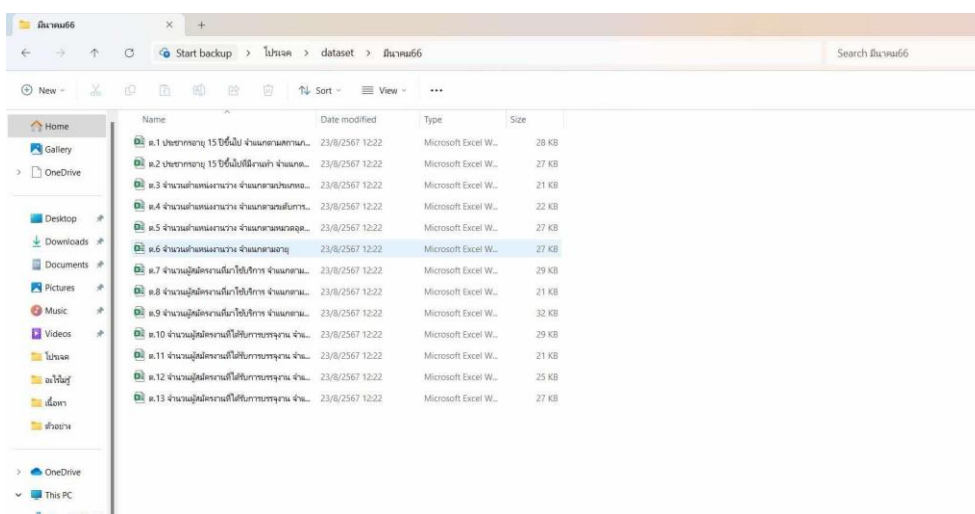
ภาพที่ 3.4 แสดงข้อมูลจำนวนกำลังแรงงาน

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาและเลือกไว้ ให้อยู่ในรูปแบบที่พร้อมสำหรับนำไป วิเคราะห์ ในขั้นตอนต่อไปได้ โดยการทำให้เป็นข้อมูลที่ถูกต้อง (Data cleaning) มักใช้เวลา ค่อนข้างมาก ระบบการรับข้อมูลป้อนเข้าสู่ระบบที่ทันสมัยในปัจจุบันจะลดการคีย์ข้อมูลจากคน ให้น้อยที่สุด แต่ จะใช้วิธีการสแกน และเลือกข้อมูล เพื่อลดความผิดพลาดให้น้อยที่สุด เพราะขั้นตอนใช้เวลา มากกว่า 50% ของเวลารวมทั้งหมด การลดข้อผิดพลาดของข้อมูลได้มากเท่าใด ก็จะมี ประสิทธิภาพมากขึ้นเท่านั้น

3.1.3 การเตรียมข้อมูล (Data Preparation)

ขั้นตอนการแปลงข้อมูลที่ได้รวบรวมมาซึ่ง เป็นขั้นตอนที่ใช้เวลามากผู้วิเคราะห์ข้อมูล ทำการจัดการกับข้อมูลที่ได้โดยรวบรวมข้อมูลและ แปลงข้อมูลที่ได้ให้อยู่ในรูปแบบที่พร้อมสำหรับ นำไปวิเคราะห์ในขั้นตอนต่อไปได้เพื่อลดความ ผิดพลาดให้น้อยที่สุด โดยใช้กระบวนการ data cleaning เพื่อให้ได้ข้อมูลที่มีประสิทธิภาพอยู่ใน รูปแบบและมาตรฐานเดียวกัน และกำจัดข้อมูลเสีย ออกดำเนินการตามขั้นตอนดังนี้

1) การปรับมาตรฐาน (Standardizing) ผู้วิเคราะห์ข้อมูลทำการรวบรวมแปลง ไฟล์ข้อมูลให้เป็นไฟล์เดียวกันอยู่รูปแบบเดียวกันพร้อมทั้งสร้างแอตทริบิวต์เพื่อกำหนดคีย์หลัก ของแต่ละไฟล์ขึ้นมาเพื่อเป็นตัวที่ใช้ระบุความสัมพันธ์ของข้อมูลที่จะแสดง



ภาพที่ 3.5 ไฟล์ทั้งหมดที่ทำการดาวน์โหลด

No.	ปีพศ	เดือน	จังหวัด	รวม	ผู้มีงานทำ	ผู้ว่างงาน	ผู้ที่รอฤดูกาล
1	2564	มกราคม	กรุงเทพมหานคร	7,651,377	5,398,074	5,287,678	110,396
2	2564	มกราคม	สมุทรปราการ	1,940,237	1,382,296	1,350,008	32,288
3	2564	มกราคม	นนทบุรี	1,462,103	980,645	958,554	22,090
4	2564	มกราคม	ปทุมธานี	1,381,184	976,810	946,385	30,425
5	2564	มกราคม	พระนครศรีอยุธยา	741,731	499,930	490,667	9,263
6	2564	มกราคม	อ่างทอง	219,200	136,667	132,714	3,953
7	2564	มกราคม	ลพบุรี	651,462	448,068	432,001	16,067
8	2564	มกราคม	สิงห์บุรี	176,465	111,051	109,865	1,186
9	2564	มกราคม	ชัยนาท	267,672	186,561	180,565	5,996
10	2564	มกราคม	สระบุรี	596,958	404,449	391,220	13,229
11	2564	มกราคม	ชลบุรี	1,508,179	1,027,007	1,019,514	7,493
12	2564	มกราคม	ระยอง	766,865	595,195	587,276	7,919
13	2564	มกราคม	จันทบุรี	470,145	365,614	363,217	2,398
14	2564	มกราคม	ตราด	239,919	173,680	171,911	1,770
15	2564	มกราคม	ฉะเชิงเทรา	688,899	449,564	441,781	7,783
16	2564	มกราคม	ปราจีนบุรี	528,400	361,109	353,518	7,590
17	2564	มกราคม	นครนายก	243,157	153,239	152,887	351

ภาพที่ 3.6 ข้อมูลที่ปรับมาตรฐานรวมเข้าไว้ด้วยกันมีจำนวนข้อมูล 2,776 รายการ

2) การแจกแจงข้อมูล (Parsing) ผู้วิเคราะห์ข้อมูลทำการสร้างตารางใหม่ขึ้นมา โดยมีจำนวนข้อมูล 2,776 รายการ ประกอบด้วย 8 แอตทริบิวต์ ประกอบด้วย ปี พ.ศ. เดือน จังหวัด ผู้มีงานทำ ผู้ว่างงาน ผู้ที่รอฤดูกาล และ ผลรวม ดังภาพที่ 3.7

	Year	Month	Province	sum	Have work	Unemployed	Waiting for work	
1								
2								
3								
4	No.							
5	1	2564	January	Bangkok	7,651,377	5,398,074	5,287,678	110,396
6	2	2564	January	Samut Prakan	1,940,237	1,382,296	1,350,008	32,288
7	3	2564	January	Nonthaburi	1,462,103	980,645	958,554	22,090
8	4	2564	January	Pathumthani	1,381,184	976,810	946,385	30,425
9	5	2564	January	Pra Nakhon Si Ayuthaya	741,731	499,930	490,667	9,263
10	6	2564	January	Angthong	219,200	136,667	132,714	3,953
11	7	2564	January	Lopburi	651,462	448,068	432,001	16,067
12	8	2564	January	Singburi	176,465	111,051	109,865	1,186
13	9	2564	January	Chai nat	267,672	186,561	180,565	5,996
14	10	2564	January	Saraburi	596,958	404,449	391,220	13,229
15	11	2564	January	Chonburi	1,508,179	1,027,007	1,019,514	7,493
16	12	2564	January	Rayong	766,865	595,195	587,276	7,919
17	13	2564	January	Chanthaburi	470,145	365,614	363,217	2,398

ภาพที่ 3.7 แสดงการแจกแจงข้อมูล

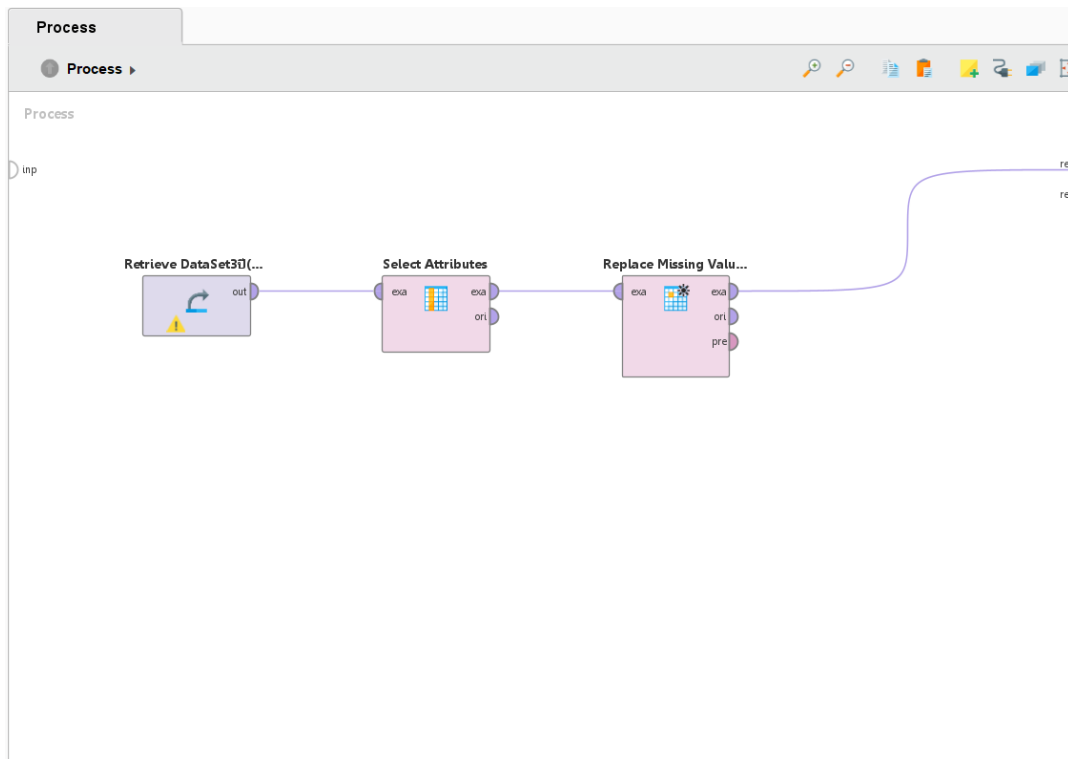
3) การแก้ไขข้อมูลที่ผิดพลาด (Correcting) ผู้วิเคราะห์ข้อมูลนำ Dataset ทำการคัดเลือกข้อมูล และทำการ Data Cleaning ข้อมูลโดยตัดส่วนที่ไม่จำเป็นออกให้เหลือเฉพาะ ข้อมูลที่จำเป็นในการวิเคราะห์ข้อมูล โดยค้นหาข้อมูลที่ผิดพลาด สูญหาย ไม่ถูกต้อง หรือยังไม่สมบูรณ์ จำนวน 284 ค่า ทำการเพิ่ม ลบ แก้ไข ข้อมูลให้ครบองค์ประกอบที่ถูกต้องเมื่อพบข้อมูลที่ไม่ถูกต้องผู้วิเคราะห์ข้อมูลจะทำการลบหรือแก้ไขข้อมูลที่พบหากพบข้อมูลที่สูญหายหรือค่าว่างผู้วิเคราะห์ ข้อมูลจะทำการตรวจสอบค้นหาข้อมูล ดังภาพที่ 3.8

No	ปีพศ	เดือน	จังหวัด	รวม	ผู้มีงานทำ	ผู้ว่างงาน	ผู้ที่รอดูฤดูกาล
2	2566	พฤษภาคม	สมุทรปราการ	1,498,812	1,478,919	19,893	-
3	2566	พฤษภาคม	นนทบุรี	1,159,428	1,144,973	14,455	-
4	2566	พฤษภาคม	ปทุมธานี	1,171,179	1,157,493	13,685	-
5	2566	พฤษภาคม	พระนครศรีอยุธยา	542,591	533,344	9,247	-
7	2566	พฤษภาคม	ลพบุรี	434,232	422,884	11,347	-
8	2566	พฤษภาคม	สิงห์บุรี	107,970	105,482	2,488	-
11	2566	พฤษภาคม	ชลบุรี	1,155,843	1,145,985	9,858	-
12	2566	พฤษภาคม	ระยอง	714,653	699,496	15,157	-
13	2566	พฤษภาคม	จันทบุรี	364,661	362,629	2,032	-
14	2566	พฤษภาคม	ตราด	174,595	173,403	1,192	-
19	2566	พฤษภาคม	ราชบุรี	479,725	475,877	3,849	-
21	2566	พฤษภาคม	สุพรรณบุรี	490,343	487,916	2,427	-
23	2566	พฤษภาคม	สมุทรสาคร	782,339	776,729	5,610	-
24	2566	พฤษภาคม	สมุทรสงคราม	102,189	101,689	501	-

ภาพที่ 3.8 แสดงข้อมูลไม่ถูกต้องที่ตรวจสอบพบ

No	ปีพศ	เดือน	จังหวัด	รวม	ผู้มีงานทำ	ผู้ว่างงาน	ผู้ที่รอดูฤดูกาล
2	2566	พฤศจิกายน	สมุทรปราการ	1,519,291	1,506,480	12,811	n.a.
3	2566	พฤศจิกายน	นนทบุรี	1,183,742	1,170,279	13,462	n.a.
4	2566	พฤศจิกายน	ปทุมธานี	1,160,068	1,136,472	23,596	n.a.
5	2566	พฤศจิกายน	พระนครศรีอยุธยา	525,621	514,895	10,726	n.a.
6	2566	พฤศจิกายน	อ่างทอง	127,818	126,936	883	n.a.
7	2566	พฤศจิกายน	ลพบุรี	459,074	453,365	5,709	n.a.
8	2566	พฤศจิกายน	สิงห์บุรี	107,228	104,902	2,326	n.a.
10	2566	พฤศจิกายน	สระบุรี	424,582	420,184	4,399	n.a.
11	2566	พฤศจิกายน	ชลบุรี	1,196,537	1,187,709	8,827	n.a.
14	2566	พฤศจิกายน	ตราด	173,892	172,609	1,284	n.a.
15	2566	พฤศจิกายน	ฉะเชิงเทรา	557,718	553,739	3,979	n.a.
16	2566	พฤศจิกายน	ปราจีนบุรี	395,764	391,886	3,878	n.a.
17	2566	พฤศจิกายน	นครนายก	148,485	147,798	687	n.a.
19	2566	พฤศจิกายน	ราชบุรี	485,437	475,831	9,607	n.a.
21	2566	พฤศจิกายน	สุพรรณบุรี	482,257	479,985	2,272	n.a.
23	2566	พฤศจิกายน	สมุทรสาคร	781,295	776,206	5,089	n.a.
24	2566	พฤศจิกายน	สมุทรสงคราม	101,557	101,063	494	n.a.
25	2566	พฤศจิกายน	เพชรบุรี	295,439	293,268	2,171	n.a.
26	2566	พฤศจิกายน	ประจวบคีรีขันธ์	282,751	281,571	1,180	n.a.

ภาพที่ 3.9 แสดงข้อมูลที่ยังไม่สมบูรณ์และสูญหาย



ภาพที่ 3.10 แสดงการนำข้อมูลเข้าสู่กระบวนการ Data Cleaning

Name	Type	Missing	Statistics			Filter (8 / 8 attributes): <input type="text" value="Search for Attributes"/>
▼ No.	Integer	0	Min 1	Max 77	Average 39	
▼ Year	Integer	0	Min 2564	Max 2566	Average 2565	
▼ Month	Nominal	0	Least September (231)	Most April (231)	Values April (231), August (231), ...[10 more]	
▼ Province	Nominal	0	Least Trat [...] (36)	Most Angtho [...] (36)	Values Angtho [...] (36), Chai n [...]	
▼ sum	Integer	0	Min 99353	Max 7651377	Average 519124.739	
▼ Have work	Integer	0	Min 97076	Max 5755751	Average 501462.817	
▼ Unemployed	Integer	8	Min 0	Max 5287678	Average 21245.180	
▼ Waiting for work	Integer	284	Min 0	Max 110396	Average 2763.091	

ภาพที่ 3.11 แสดงค่า Missing Value ที่ตรวจพบ

ExampleSet (Replace Missing Values) x

Open in Turbo Prep Auto Model Interactive Analysis Filter (2,772 / 2,772 example)

Row No.	Unemployed	Waiting ... ↑	No.	Year	Month	Province	sum	Have work
47	579284	0	47	2564	January	Sisaket ...	811359	579284
79	32288	0	2	2564	February	Samut Prakan ...	1382296	1350008
80	22090	0	3	2564	February	Nonthaburi ...	980645	958554
81	30425	0	4	2564	February	Pathumthani ...	976810	946385
85	1186	0	8	2564	February	Singburi ...	111051	109865
86	5996	0	9	2564	February	Chai nat ...	186561	180565
87	13229	0	10	2564	February	Saraburi ...	404449	391220
88	7493	0	11	2564	February	Chonburi ...	1027007	1019514
90	2398	0	13	2564	February	Chanthaburi ...	365614	363217
91	1770	0	14	2564	February	Trat ...	173680	171911
92	7783	0	15	2564	February	Chachoengsao...	449564	441781
99	8298	0	22	2564	February	Nakhon Patho...	691242	682944
100	10264	0	23	2564	February	Samut Sakhon...	719882	709617
101	1220	0	24	2564	February	Samut Songkh...	111435	110215
102	3552	0	25	2564	February	Phetchaburi ...	282548	278995
104	20498	0	27	2564	February	Chiang Mai ...	1009845	989347
105	2102	0	28	2564	February	Lamphun ...	264670	262568
106	3535	0	29	2564	February	Lampang ...	407383	403848

ExampleSet (2,772 examples,0 special attributes,8 regular attributes)

ภาพที่ 3.12 แสดงข้อมูลที่สมบูรณ

4) การจัดหมวดหมู่ (Transform) ผู้วิเคราะห์ข้อมูลนำ Dataset ทำการคัดเลือกข้อมูล และทำการ Data Cleaning มาทำการจัดหมวดหมู่เพื่อความถูกต้อง ผู้วิเคราะห์ข้อมูลดำเนินการจัดกลุ่มแบ่งข้อมูล 3 แอททริบิวต์ คือแอททริบิวต์ ผู้มีงานทำ(Have work) ผู้ว่างงาน(Unemployee) และผู้ที่รอฤดูกาล(Waiting for Work) เพื่อแบ่งข้อมูลอย่างชัดเจน เตรียมสู่การนำข้อมูลไปวิเคราะห์โดยดำเนินการตามขั้นตอน ดังภาพที่ 3.14

จังหวัด	ประชากร อายุ 15 ปี ขึ้นไป	กำลังแรงงานรวม				ผู้ไม่อยู่ในกำลังแรงงาน			
		รวม	ผู้มีงานทำ	ผู้ว่างงาน	ผู้ที่รอฤดูกาล	รวม	ทำงานบ้าน	เรียนหนังสือ	อื่น ๆ
รวม	56,940,473	39,096,342	38,289,037	727,048	70,463	17,705,004	4,863,715	4,187,166	8,804,255
1. กรุงเทพมหานคร	7,654,042	5,482,740	5,349,947	132,053	740	2,171,302	571,471	496,502	1,103,329
2. สมุทรปราการ	1,948,344	1,406,474	1,371,893	34,581	0	541,870	161,541	100,576	279,753
3. นนทบุรี	1,469,435	1,000,059	972,952	27,107	0	469,376	132,648	89,890	246,837
4. ปทุมธานี	1,386,459	962,344	941,601	20,743	0	424,115	146,919	85,658	191,538
5. พระนครศรีอยุธยา	741,840	517,238	506,969	10,269	0	224,602	53,235	54,764	116,604
6. อ่างทอง	219,435	143,644	140,988	2,656	0	75,791	19,605	14,836	41,350
7. ลพบุรี	651,758	474,037	458,182	11,224	4,631	177,721	36,807	44,302	96,612
8. สิงห์บุรี	176,680	113,348	112,105	1,120	123	63,332	18,546	10,806	33,980
9. ชัยนาท	267,993	182,781	176,526	6,255	0	85,212	18,607	14,384	52,221
10. สระบุรี	597,021	408,239	394,647	13,153	439	188,782	57,363	36,186	95,230
11. ชลบุรี	1,511,544	1,060,008	1,047,022	12,986	0	451,536	152,278	97,797	201,462
12. ระยอง	769,002	589,044	583,242	5,427	375	179,958	65,446	45,354	69,157
13. จันทบุรี	471,967	372,273	368,295	3,710	267	99,694	22,945	26,465	50,284
14. ตราด	240,743	175,258	173,661	1,597	0	65,485	22,673	11,920	30,892
15. ฉะเชิงเทรา	691,495	449,500	443,677	5,823	0	241,995	105,093	46,875	90,027
16. ปราจีนบุรี	530,424	373,868	364,623	9,245	0	156,556	43,492	33,819	79,245

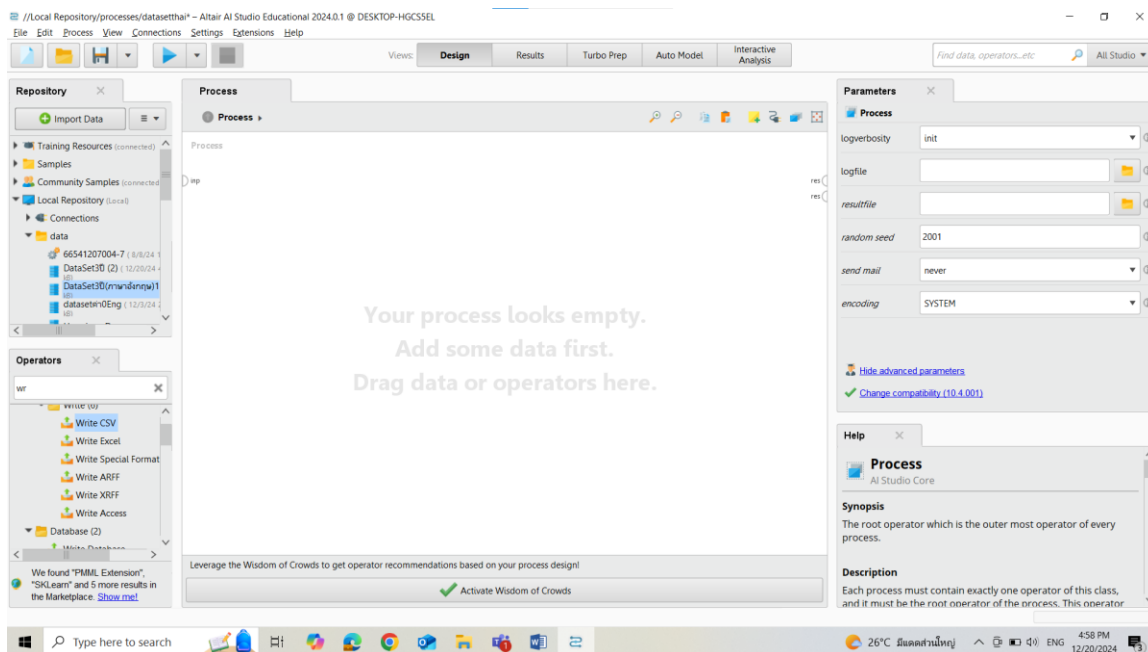
ภาพที่ 3.13 แสดงชุดข้อมูลที่จะต้องการจัดกลุ่ม

No.	ปีพศ	เดือน	จังหวัด	รวม	ผู้มีงานทำ	ผู้ว่างงาน	ผู้ที่รอฤดูกาล
1	2564	มกราคม	กรุงเทพมหานคร	7,651,377	5,398,074	5,287,678	110,396
2	2564	มกราคม	สมุทรปราการ	1,940,237	1,382,296	1,350,008	32,288
3	2564	มกราคม	นนทบุรี	1,462,103	980,645	958,554	22,090
4	2564	มกราคม	ปทุมธานี	1,381,184	976,810	946,385	30,425
5	2564	มกราคม	พระนครศรีอยุธยา	741,731	499,930	490,667	9,263
6	2564	มกราคม	อ่างทอง	219,200	136,667	132,714	3,953
7	2564	มกราคม	ลพบุรี	651,462	448,068	452,001	16,067
8	2564	มกราคม	สิงห์บุรี	176,465	111,051	109,865	1,186
9	2564	มกราคม	ชัยนาท	267,672	186,561	180,565	5,996
10	2564	มกราคม	สระบุรี	596,958	404,449	391,220	13,229
11	2564	มกราคม	ชลบุรี	1,508,179	1,027,007	1,019,514	7,493
12	2564	มกราคม	ระยอง	766,865	595,195	587,276	7,919
13	2564	มกราคม	จันทบุรี	470,145	365,614	363,217	2,398
14	2564	มกราคม	ตราด	239,919	173,680	171,911	1,770
15	2564	มกราคม	ฉะเชิงเทรา	688,899	449,564	441,781	7,783
16	2564	มกราคม	ปราจีนบุรี	528,400	361,109	353,518	7,590

ภาพที่ 3.14 แสดงการจัดการจัดหมวดหมู่

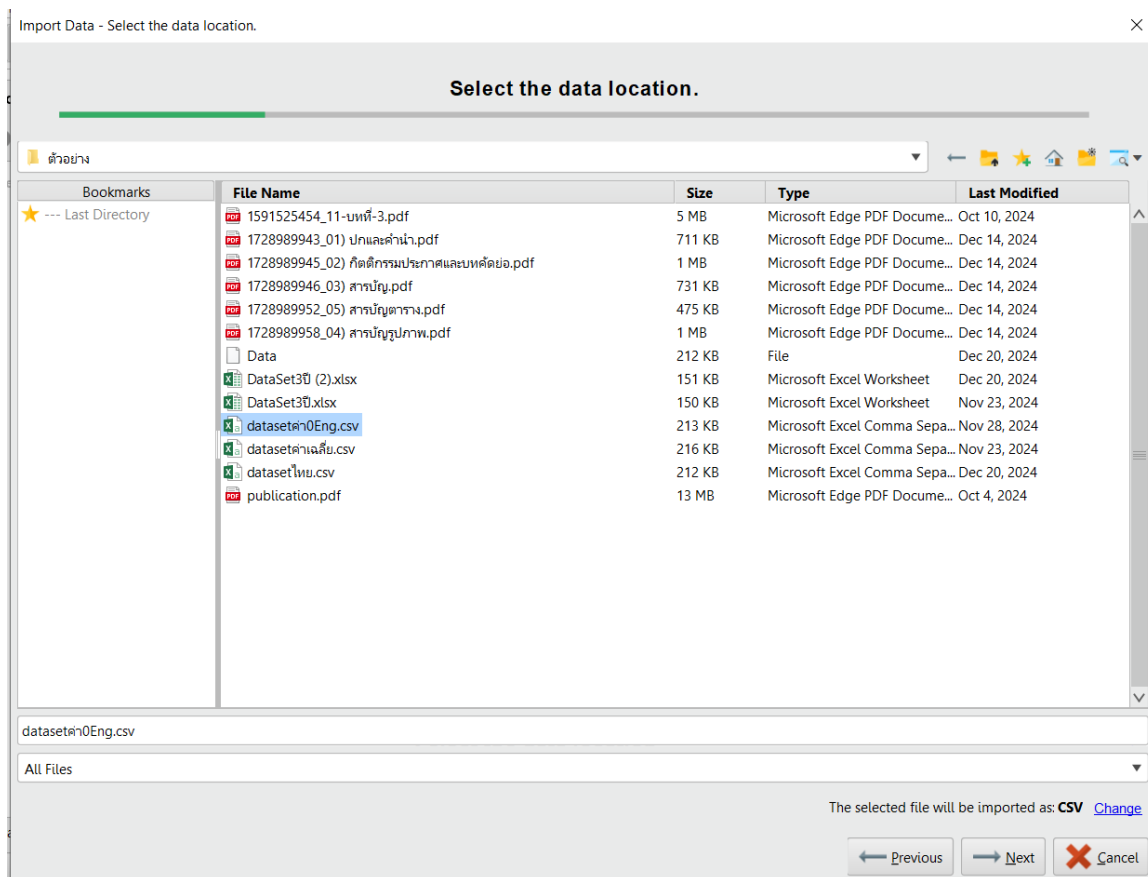
5) การโหลดข้อมูล (Loading) เข้าสู่โปรแกรม Rapid miner Studio เพื่อจะนำไปวิเคราะห์และสร้างออกมาเป็นรายงานภาพ Visualization

5.1 โหลดข้อมูลที่ได้ทำการเตรียมพร้อมไว้แล้วโดยใช้โปรแกรม Rapid miner



ภาพที่ 3.15 แสดงหน้าต่างการทำงานพื้นที่ของโปรแกรม Rapid miner Studio

5.2 ใช้ฟังก์ชัน Import Data ในการนำเข้าข้อมูลเพื่อนำมาวิเคราะห์เลือกชุดข้อมูลที่จัดเตรียมแล้วกด Next



ภาพที่ 3.16 แสดงหน้า Select the Data location

5.3 ทำการตรวจสอบข้อมูลว่าถูกต้องหรือไม่ ถ้าข้อมูลถูกต้องแล้วกด Next

Import Data - Specify your data format

Specify your data format

Header Row
 Start Row
 Column Separator

File Encoding
 Escape Character
 Decimal Character

Use Quotes
 Skip Comments
 Trim Lines Multiline Text

1	Unemployed	Waiting for work	No.	Year	Month	Province	sum	Have work
2	5287678	110396	1	2564	January	Bangkok	7651377	976810
3	1350008	32288	2	2564	January	Samut Prakan	1940237	1382296
4	958554	22090	3	2564	January	Nonthaburi	1462103	980645
5	946385	30425	4	2564	January	Pathumthani	1381184	976810
6	490667	9263	5	2564	January	Pra NaKhon Si Ay...	741731	499930
7	132714	3953	6	2564	January	Angthong	219200	136667
8	432001	16067	7	2564	January	Lopburi	651462	448068
9	109865	1186	8	2564	January	Singburi	176465	111051
10	180565	5996	9	2564	January	Chai nat	267672	186561
11	391220	13229	10	2564	January	Saraburi	596958	404449
12	1019514	7493	11	2564	January	Chonburi	1508179	1027007
13	587276	7919	12	2564	January	Rayong	766865	595195
14	363217	2398	13	2564	January	Chanthaburi	470145	365614
15	171911	1770	14	2564	January	Trat	239919	173680
16	441781	7783	15	2564	January	Chachoengsao	688899	449564
17	353518	7590	16	2564	January	Prachinburi	528400	361109
18	152887	351	17	2564	January	Nakhon Navok	243157	153239

no problems.

ภาพที่ 3.17 แสดงหน้า Specify your data format

5.4 ตรวจสอบชนิดข้อมูล หากผิดพลาดสามารถแก้ไขข้อมูลในหน้านี้ให้เสร็จสมบูรณ์ก่อนนำข้อมูลเข้าสู่โปรแกรม เพื่อผลการวิเคราะห์ที่แม่นยำและถูกต้อง

Import Data - Format your columns. ×

Format your columns.

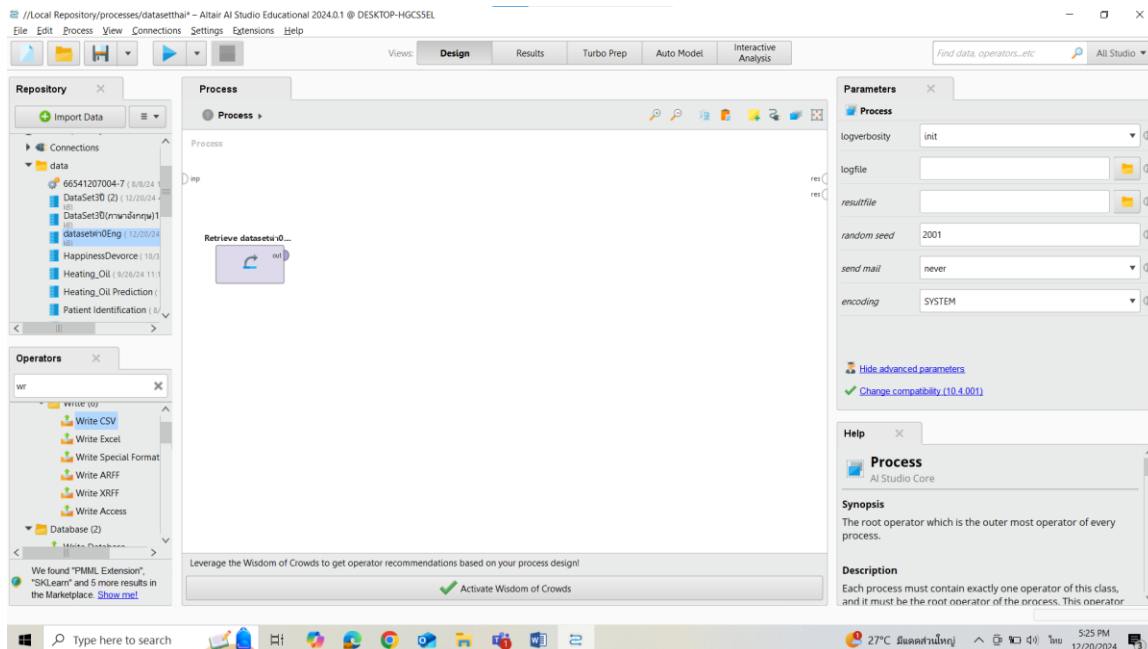
Date format Replace errors with missing values ⓘ

	Unemployed <i>integer</i>	Waiting for ... <i>integer</i>	No. <i>integer</i>	Year <i>integer</i>	Month <i>polynomial</i>	Province <i>polynomial</i>	sum <i>integer</i>	Have work <i>integer</i>
1	5287678	110396	1	2564	January	Bangkok	7651377	976810
2	1350008	32288	2	2564	January	Samut Prakan	1940237	1382296
3	958554	22090	3	2564	January	Nonthaburi	1462103	980645
4	946385	30425	4	2564	January	Pathumthani	1381184	976810
5	490667	9263	5	2564	January	Pra Nakhon Si Ayu...	741731	499930
6	132714	3953	6	2564	January	Anghong	219200	136667
7	432001	16067	7	2564	January	Lopburi	651462	448068
8	109865	1186	8	2564	January	Singburi	176465	111051
9	180565	5996	9	2564	January	Chai nat	267672	186561
10	391220	13229	10	2564	January	Saraburi	596958	404449
11	1019514	7493	11	2564	January	Chonburi	1508179	1027007
12	587276	7919	12	2564	January	Rayong	766865	595195
13	363217	2398	13	2564	January	Chanthaburi	470145	365614
14	171911	1770	14	2564	January	Trat	239919	173680
15	441781	7783	15	2564	January	Chachoengsao	688899	449564
16	353518	7590	16	2564	January	Prachinburi	528400	361109
17	152887	351	17	2564	January	Nakhon Nayok	243157	153239
18	309083	13469	18	2564	Januarv	Sa kaeo	504671	322552

✔ no problems.

ภาพที่ 3.18 แสดงหน้า format your columns

5.5 เมื่อโหลดข้อมูลสำเร็จ ข้อมูลจะแสดงดังภาพที่ 3.19



ภาพที่ 3.19 แสดงการนำข้อมูลเข้าสู่โปรแกรม Rapid miner Studio

3.1.4 การสร้างแบบจำลอง (Modeling) คณะผู้วิจัยวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูลแบบ Data Classification เพื่อการเปรียบเทียบประสิทธิภาพโมเดลเพื่อการพยากรณ์อัตราการว่างงานในประเทศไทยในช่วงปี 2564-2566 เป็นเทคนิคที่ต้องใช้ข้อมูลในอดีตมาสร้างรูปแบบการพยากรณ์ ในรูปของสมการคณิตศาสตร์ใช้วิธีการในรูปแบบของสมการทาง คณิตศาสตร์ ซึ่งผู้วิเคราะห์ข้อมูลได้เลือกใช้เทคนิครูปแบบอนุกรมเวลา (Time Series Models) ซึ่งได้แก่ ARIMA (Auto Regressive Integrated Moving Average) , Holt's winter และ K-NN โดยใช้ โปรแกรมที่ใช้ทำเหมืองข้อมูล ด้วยชุดข้อมูลที่คัดเลือกโดยมาขั้นตอนการดำเนินการและรายละเอียด

ในการพยากรณ์โดยใช้รูปแบบอนุกรมเวลาจะเป็นการพยากรณ์ข้อมูลในอนาคตจากข้อมูล ในอดีตเท่านั้น ตัวแปรอื่น ๆ จะไม่นำมาพิจารณา ซึ่งรูปแบบของข้อมูล (Data Pattern) มีเทคนิคการ พยากรณ์หลายรูปแบบ แต่โดยปกติข้อมูลจะแบ่งได้หลายรูปแบบซึ่งทาง ผู้วิเคราะห์ข้อมูลได้เลือกข้อมูลในรูปแบบมี ความต้องการเปลี่ยนไปอย่างมี ทิศทาง

Unemployed	Waiting for ...	No.	Year	Month	Province	sum	Have work
5287678	110396	1	2564	January	Bangkok	7651377	976810
1350008	32288	2	2564	January	Samut Prakan	1940237	1382296
958554	22090	3	2564	January	Nonthaburi	1462103	980645
946385	30425	4	2564	January	Pathumthani	1381184	976810
490667	9263	5	2564	January	Pra NaKhon Si ...	741731	499930
132714	3953	6	2564	January	Angthong	219200	136667
432001	16067	7	2564	January	Lopburi	651462	448068
109865	1186	8	2564	January	Singburi	176465	111051
180565	5996	9	2564	January	Chai nat	267672	186561
391220	13229	10	2564	January	Saraburi	596958	404449
1019514	7493	11	2564	January	Chonburi	1508179	1027007
587276	7919	12	2564	January	Rayong	766865	595195
363217	2398	13	2564	January	Chanthaburi	470145	365614
171911	1770	14	2564	January	Trat	239919	173680
441781	7783	15	2564	January	Chachoengsao	688899	449564
353518	7590	16	2564	January	Prachinburi	528400	361109
152887	351	17	2564	January	Nakhon Nayok	243157	153239
309083	13469	18	2564	January	Sa kaeo	504671	322552

ภาพที่ 3.20 แสดงตารางข้อมูลที่น่ามาใช้

จากภาพที่ 3.20 ประกอบด้วย 8 แอตทริบิวต์ ดังตารางที่ 3.1

ตารางที่ 3.1 แสดงตารางข้อมูลที่นำมาใช้ รายละเอียดของแอตทริบิวต์

ชื่อแอตทริบิวต์	รายละเอียด
ปี พ.ศ.	ปีพ.ศ.ที่นำข้อมูลมาวิเคราะห์ ตั้งแต่ พ.ศ. 2564-2566
เดือน	เดือนที่นำมาวิเคราะห์ 12 เดือน
จังหวัด	จังหวัดที่นำข้อมูลมาวิเคราะห์ 77 จังหวัด
ผู้ปฏิบัติงาน	บุคคลที่อยู่ในกำลังแรงงานและมีสถานะทำงานหรือประกอบอาชีพในช่วงเวลาที่กำหนด
ผู้ว่างงาน	บุคคลที่อยู่ในช่วงอายุและสถานะที่สามารถทำงานได้ แต่ในขณะสำรวจข้อมูลยังไม่มีการทำงาน
ผู้ที่รอฤดูกาล	บุคคลที่อยู่ในช่วงอายุที่สามารถทำงานได้ และเคยมีงานทำมาก่อน แต่ในช่วงเวลาที่สำรวจข้อมูล ไม่มีการทำงานเนื่องจากลักษณะงานของพวกเขาขึ้นอยู่กับฤดูกาลหรือช่วงเวลาที่กำหนด
รวม	ผลรวมของจำนวนผู้ปฏิบัติงาน ผู้ว่างงาน และผู้ที่รอฤดูกาล

3.1.4.1 ขั้นตอนการพยากรณ์อัตราการว่างงานโดยใช้โมเดล ARIMA

ส่วนประกอบของสมการ ARIMA AR (Autoregressive): ส่วนนี้แสดงถึงการพึ่งพาของค่าปัจจุบันกับค่าที่ผ่านมา (Lagged values) ซึ่งสามารถเขียนได้ในรูปสมการ

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

โดยที่

Y_t ค่าของตัวแปร ณ เวลา t

ϕ_i ค่าสัมประสิทธิ์ AR

p จำนวน lag

ϵ_t ค่า error หรือ residual

I (Integrated): ส่วนนี้ใช้เพื่อให้ข้อมูลอนุกรมเวลามีความนิ่ง (Stationary) โดยการนำค่าความต่าง (Difference) มาใช้

$$Y_t' = Y_t - Y_{t-1}$$

โดยที่

Y_t' ค่าความต่างอันดับที่ 1 (First-order difference)

MA (Moving Average): ส่วนนี้แสดงถึงการพึ่งพาของค่าปัจจุบันกับค่า residual จาก lag ที่ผ่านมา

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

โดยที่

ϵ_t ค่า error ณ เวลา t

θ_i ค่าสัมประสิทธิ์ MA

q จำนวน lag ของ residual

สมการ ARIMA ที่รวมทุกส่วนเข้าด้วยกัน

$$\Phi_p(B)(1-B)^d Y_t = \Theta_q(B)\epsilon_t$$

โดยที่

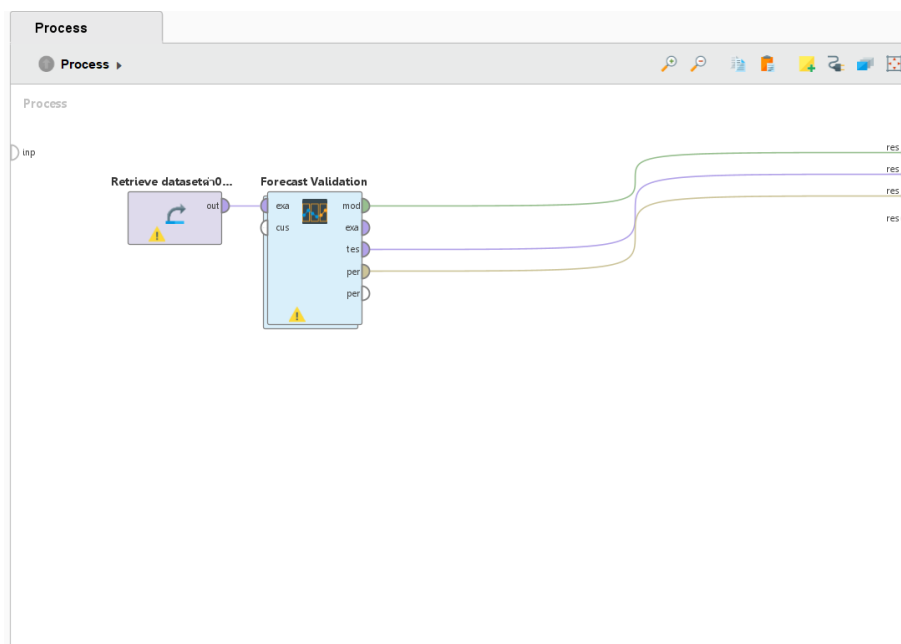
$\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ส่วน AR

$\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ ส่วน MA

$(1-B)^d$ การทำ Differencing

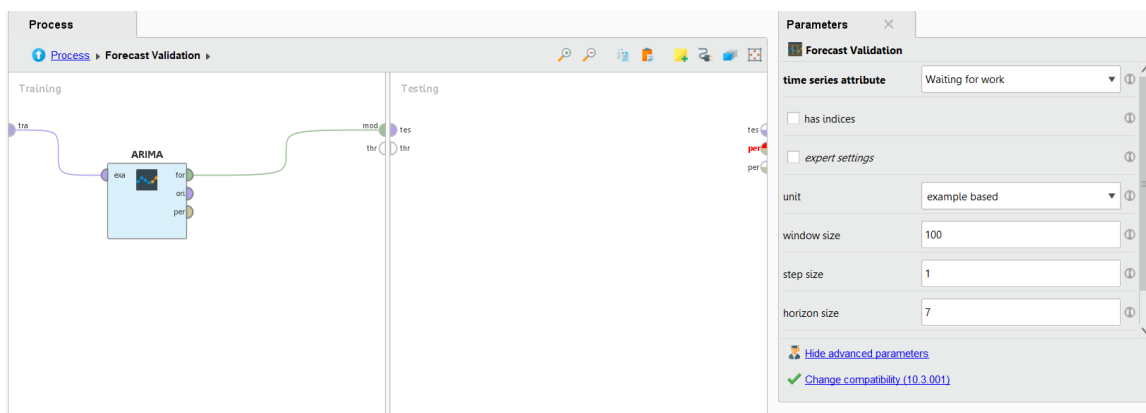
B Backshift operator ($BY_t = Y_{t-1}$)

ขั้นตอนที่ 1 เลือกใช้ Operator ที่ชื่อว่า forecast validation



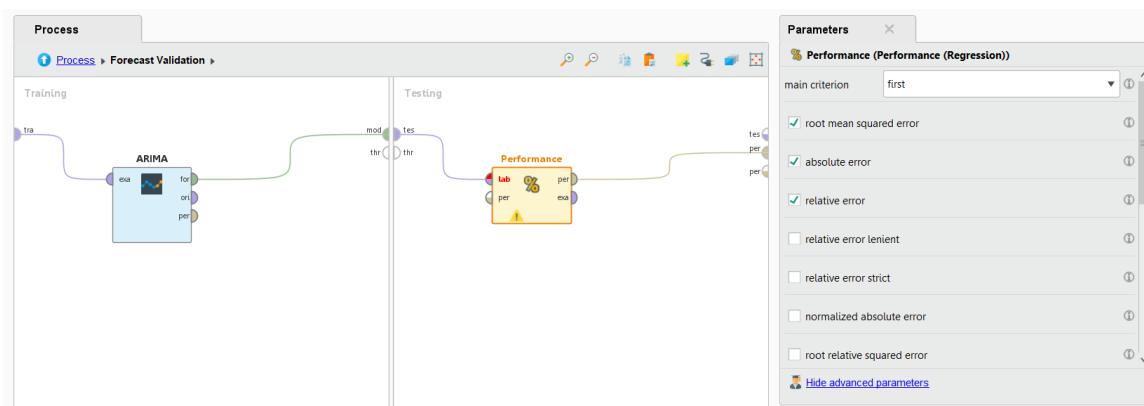
ภาพที่ 3.21 แสดงขั้นตอนที่ 1 ของการพยากรณ์

ขั้นตอนที่ 2 เลือกใช้ Operator ที่ชื่อว่า ARIMA



ภาพที่ 3.22 แสดงขั้นตอนที่ 2 ของการพยากรณ์

ขั้นตอนที่ 3 เลือกใช้ Operator ที่ชื่อว่า Performance ในการประเมินประสิทธิภาพของโมเดลที่เลือกใช้



ภาพที่ 3.23 แสดงขั้นตอนที่ 3 ของการพยากรณ์

ขั้นตอนที่ 4 ผลลัพธ์โมเดล ARIMA

% PerformanceVector (Performance) ✕
■ ArimaModel (ARIMA) ✕

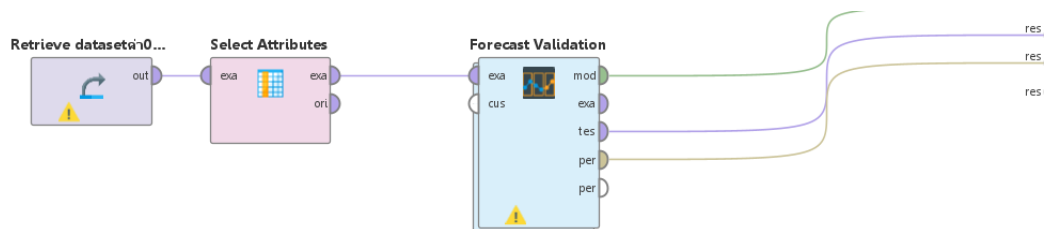
ArimaModel

Forecast Model trained on the following time series:
 Name of time series: Waiting for work Number of values: 100

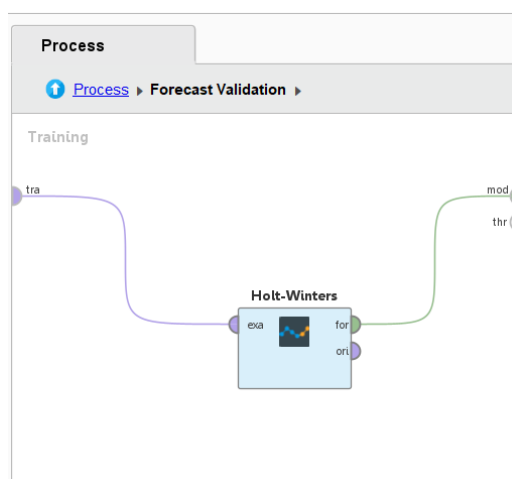
Resulting Forecast Model:
 Arima Model (p: 1,d: 0, q: 1)
 AR Coefficients: [-0.44519235130930684]
 MA Coefficients: [0.4524881976105697]
 constant: 528.8890276735223, length of residuals: 1

ภาพที่ 3.24 แสดงขั้นตอนที่ 4 ของการพยากรณ์

3.1.4.2 ขั้นตอนการพยากรณ์อัตราการว่างงานโดยใช้โมเดล Holt's winter
 ขั้นตอนที่ 1 เลือกใช้ Operator ที่ชื่อว่า forecast validation

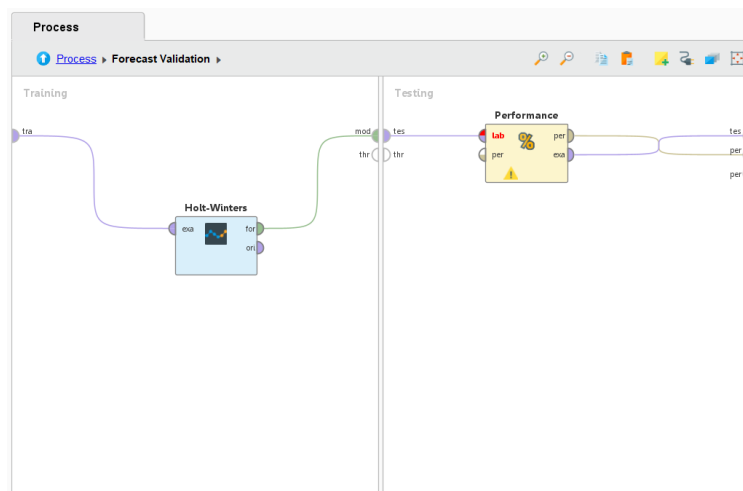


ภาพที่ 3.25 แสดงขั้นตอนที่ 1 ของการพยากรณ์โดยใช้โมเดล Holt's winter
 ขั้นตอนที่ 2 เลือกใช้ Operator ที่ชื่อว่า Holt's winter ทำการวิเคราะห์
 ข้อมูลโดยใช้แอทริบิวต์ ผู้ว่างงาน(Unemployee)



ภาพที่ 3.26 แสดงขั้นตอนที่ 2 ของการพยากรณ์โดยใช้โมเดล Holt's winter

ขั้นตอนที่ 3 เลือกใช้ Operator ที่ชื่อว่า Performance ในการประเมินประสิทธิภาพของโมเดลที่เลือกใช้



ภาพที่ 3.27 แสดงขั้นตอนที่ 3 ของการพยากรณ์โดยใช้โมเดล Holt's winter

ขั้นตอนที่ 4 ผลลัพธ์โมเดล Holt's Winters Model

HoltWintersModel

Forecast Model trained on the following time series:

Name of time series: Waiting for work Number of values: 100

Resulting Forecast Model:

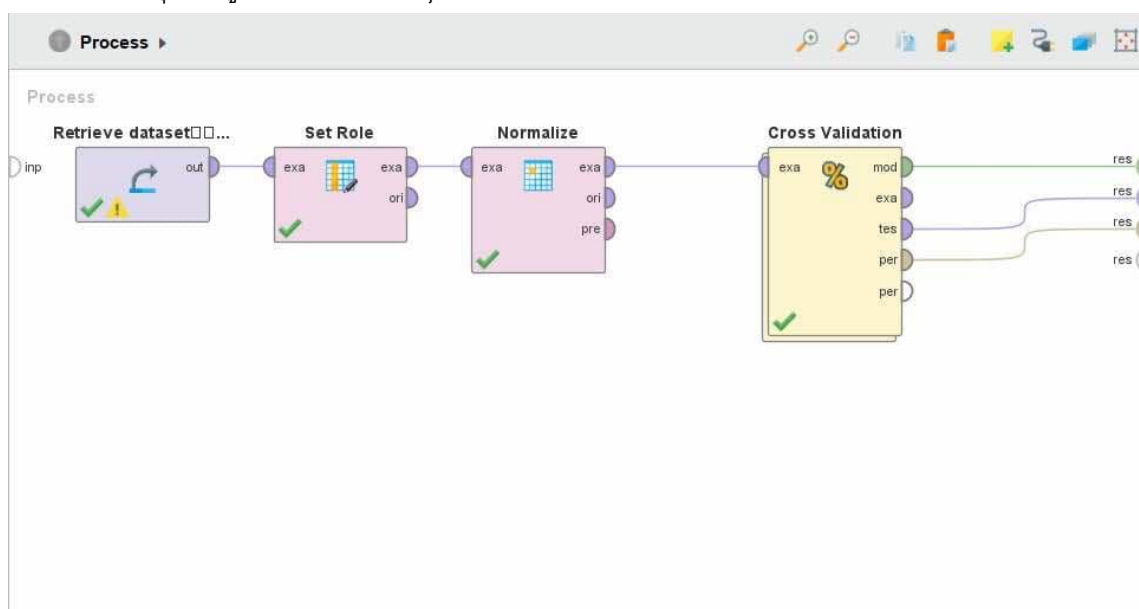
Holt-Winters Model (alpha: 0.5, beta: 0.1, gamma: 0.5)

Period: 3, mode type: ADDITIVE

ภาพที่ 3.28 แสดงผลลัพธ์ ของการพยากรณ์โดยใช้โมเดล Holt's winter

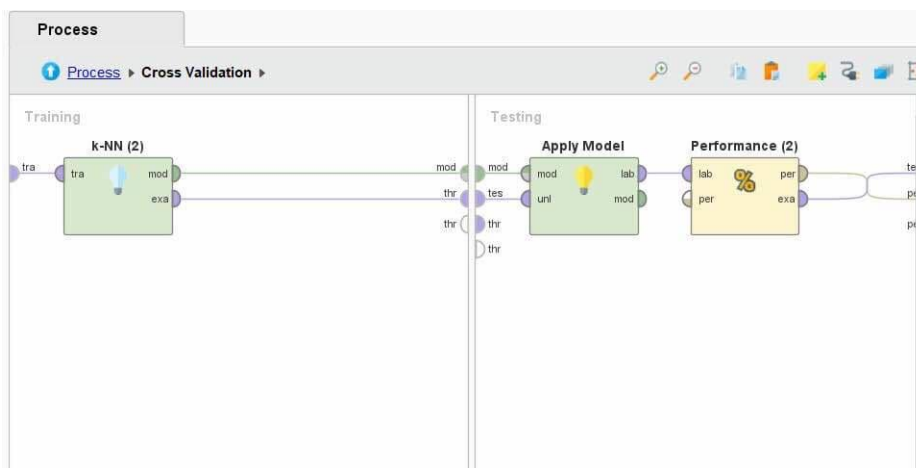
3.1.4.3 ขั้นตอนการพยากรณ์อัตราการว่างงานโดยใช้โมเดล k-NN

ขั้นตอนที่ 1 ใช้ Operator ที่ชื่อว่า Set Role กำหนด คอลัมน์ Unemployed กำหนดให้เป็น Label ซึ่งคือเป็นตัวแปรเป้าหมาย (output) คอลัมน์ Year, Month, Province, sum, Waiting for work, และ Have work ถูกกำหนดเป็น "Regular" เพื่อใช้เป็นตัวแปรต้น Normalize ทำการปรับข้อมูลให้มีขนาดหรือช่วงค่าที่เหมาะสม การปรับให้ค่าต่าง ๆ อยู่ในช่วง 0 ถึง 1 หรือ -1 ถึง 1 เพื่อให้โมเดลทำงานได้ดีขึ้น Cross Validation เป็นกระบวนการตรวจสอบความแม่นยำของโมเดล โดยการแบ่งชุดข้อมูลออกเป็นส่วน ๆ



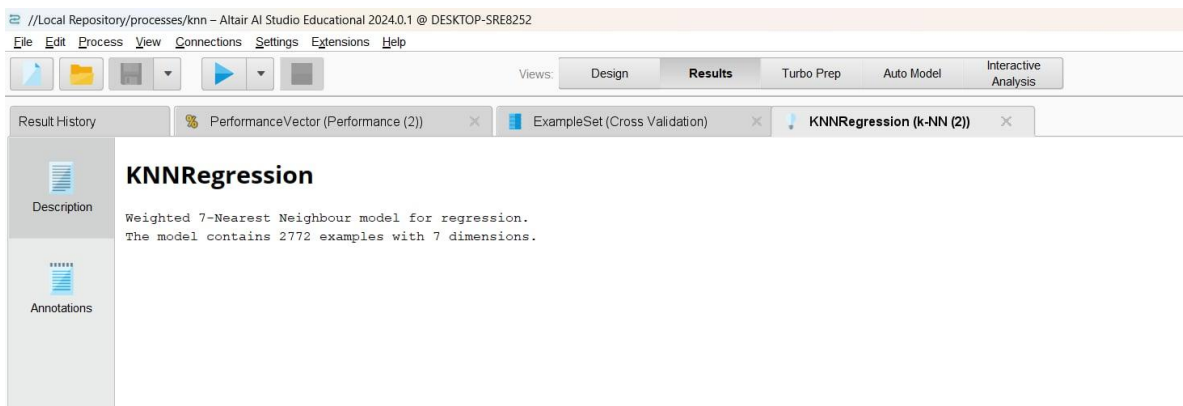
ภาพที่ 3.29 แสดงขั้นตอนที่ 1 การพยากรณ์อัตราการว่างงานโดยใช้โมเดล k-NN

ขั้นตอนที่ 2 เลือกใช้ Operator k-NN (k-Nearest Neighbors) ในการพยากรณ์ Apply Model นำมาใช้กับชุดข้อมูลการทดสอบ (test) เพื่อสร้างผลลัพธ์และประเมินประสิทธิภาพด้วย Performance



ภาพที่ 3.30 แสดงขั้นตอนที่ 2 การพยากรณ์อัตราการใช้งานโดยใช้โมเดล k-NN

ขั้นตอนที่ 3 ผลลัพธ์โมเดล k-NN (k-Nearest Neighbors)



ภาพที่ 3.31 แสดงผลลัพธ์ ของการพยากรณ์โดยใช้โมเดล Holt's winter

ค่าเฉลี่ยของร้อยละความผิดพลาดสัมบูรณ์ (Mean Absolute Percentage Error, MAPE) เป็นวิธีวัดความแม่นยำ โดยคำนวณเปอร์เซ็นต์ความผิดพลาดในการพยากรณ์ โดยไม่คำนึงถึง เครื่องหมาย ค่าที่ได้ต่ำ มีความแม่นยำสูง หากค่าความคลาดเคลื่อนต่ำที่สุดจะทำให้การ

พยากรณ์ใกล้เคียงกับความจริงมากที่สุดหมายความว่าหากค่าความคลาดเคลื่อนยิ่งน้อยยิ่ง เป็นผลดีกับโมเดล (รุ่งนภา, 2557) เช่น ถ้า MAPE = 4% แสดงว่าวิธีที่เลือกมีความคลาดเคลื่อนในการพยากรณ์ เฉลี่ยอยู่ที่ร้อยละ 4

$$MAPE = \frac{[\sum |A_t - F_t| / A_t] \times 100}{N}$$

โดย F_t หมายถึงค่าพยากรณ์ในงวดที่ 1

A_t หมายถึงยอดที่เกิดขึ้นจริงในช่วงเวลา t

N หมายถึง จำนวนข้อมูล

สามารถคำนวณค่าพยากรณ์ออกมาได้มีค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ ดังตารางที่ 3.2

ตารางที่ 3.2 แสดงค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์

Model	MAPE
ARIMA	17%
Holt's winter	6%
K-NN(K-Nearest Neighbour)	9%

3.1.5 การประเมินประสิทธิภาพ (Evaluation) เป็นขั้นตอนก่อนนำผลลัพธ์ที่ได้จากขั้นตอนที่ 4 ไปใช้งานผู้วิเคราะห์ข้อมูลวัดประสิทธิผลของผลลัพธ์ที่ได้กับวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก ว่ามีนัยสำคัญหรือความน่าเชื่อถือมากน้อยเพียงใดโดยเปรียบเทียบนำผลที่ได้ สามารถวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ โดยผู้วิเคราะห์ข้อมูลได้เปรียบเทียบนำผลที่ได้ดังนี้

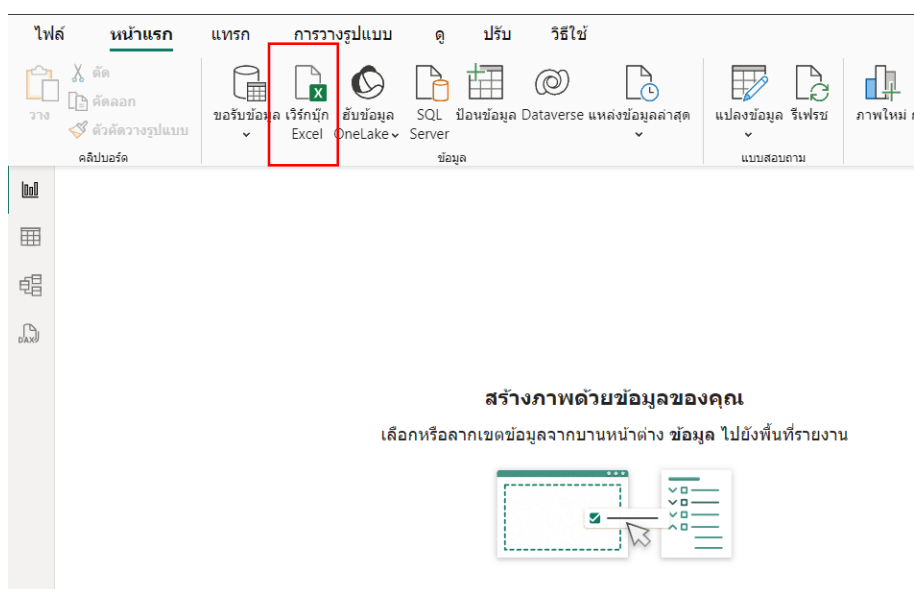
3.1.5.1 การวัดประสิทธิภาพของการวิเคราะห์อนุกรมเวลา จากการสร้างโมเดล Time Series Model โดยใช้ Operator ที่ชื่อว่า Performance ในโปรแกรม rapid miner studio เพื่อหาค่า RMSE ,MAE, Relative Error และ Squared Error ของโมเดล ARIMA, Holt's winters และ K-NN ดังตารางที่ 3.3

ตารางที่ 3.3 แสดงการวัดประสิทธิภาพของการวิเคราะห์อนุกรมเวลา

Model	RMSE	MAE	Relative Error	Squared Error
ARIMA	4231.482	3250.121	462.16%	554147953.621
Holt's winters	5101.080	4073.976	575.72%	86922612.346
K-NN(K-Nearest Neighbour)	71327.397	10770.948	109.39%	11599905173.965

3.1.6 การนำผลลัพธ์ไปใช้งาน (Deployment) ผู้วิเคราะห์ข้อมูลนำผลของค้ความรูู้ที่ได้เหล่านี้ไปนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพโดยใช้โปรแกรม Power Bi โดยดำเนินการตามขั้นตอนดังนี้

6.1 เข้ามาสู่หน้า Dash Board การทำงาน import ข้อมูลผ่าน Workbook Excel



ภาพที่ 3.32 แสดงการเข้ามาสู่หน้า Dash Board การทำงาน

6.2 เลือกWorksheet ที่จะใช้ในการทำ Dash Board แล้วกดโหลด

หน้างาน

ค้นหา

ตัวเลือกการแสดง

- MAPE.xlsx [6]
 - ARIma
 - arima2
 - Holt
 - Holt2
 - K-NN
 - K-NN2
- ตารางที่แนะนำ [4]
 - ตาราง 1 (ARIma)
 - Month (arima2)
 - Month (Holt2)
 - Month (K-NN2)

arima2

NO.	Ft	At	Month	Province
1	16545	3819.47813	January	Bangkok
2	3596	4322.815763	January	Samut Prakan
3	2419	4655.212699	January	Nonthaburi
4	2826	4624.020564	January	Pathumthani
5	5776	4694.50479	January	Pra Nakhon Si Ayuthaya
6	2351	4663.346247	January	Angthong
7	20511	4687.209543	January	Lopburi
8	1220	21107.17003	January	Singburi
9	3552	22556.87395	January	Chai nat
10	2308	22303.14205	January	Saraburi
11	20498	21739.76249	January	Chonburi
12	2102	21119.35267	January	Rayong
13	3535	20487.69696	January	Chanthaburi
14	3536	19853.09203	January	Trat
15	3552	134385.1676	January	Chachoengsao
16	2308	169999.1031	January	Prachinburi
17	20498	200421.3329	January	Nakhon Nayok
18	2102	225014.5378	January	Sa kaeo
19	3535	244924.7025	January	Ratchaburi
20	3536	261042.9292	January	Kanchanaburi
21	2530	274091.4157	January	Suphanburi
22	2308	132412.8382	January	Nakhon Pathom
23	20498	168829.623	January	Samut Sakhon

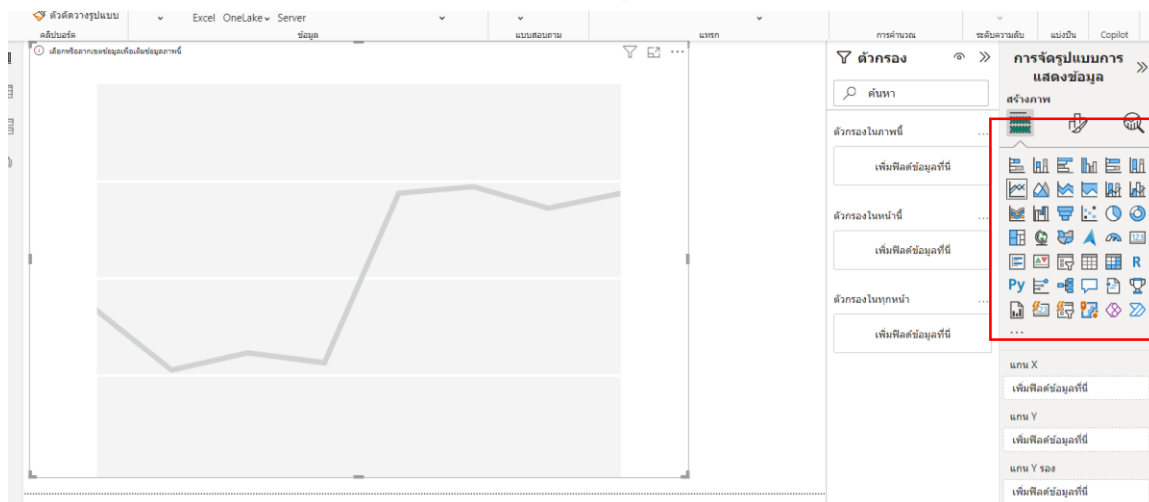
โหลด

แปลงข้อมูล

ยกเลิก

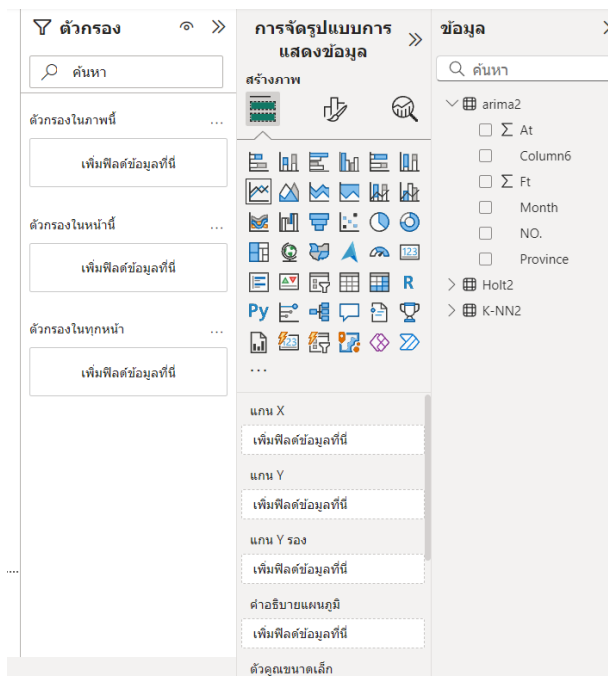
ภาพที่ 3.33 แสดงการโหลด ข้อมูลที่ผ่านการพยากรณ์แล้ว ลงในโปรแกรม

6.3 เลือกใช้กราฟที่เหมาะสมกับข้อมูลที่นำมา



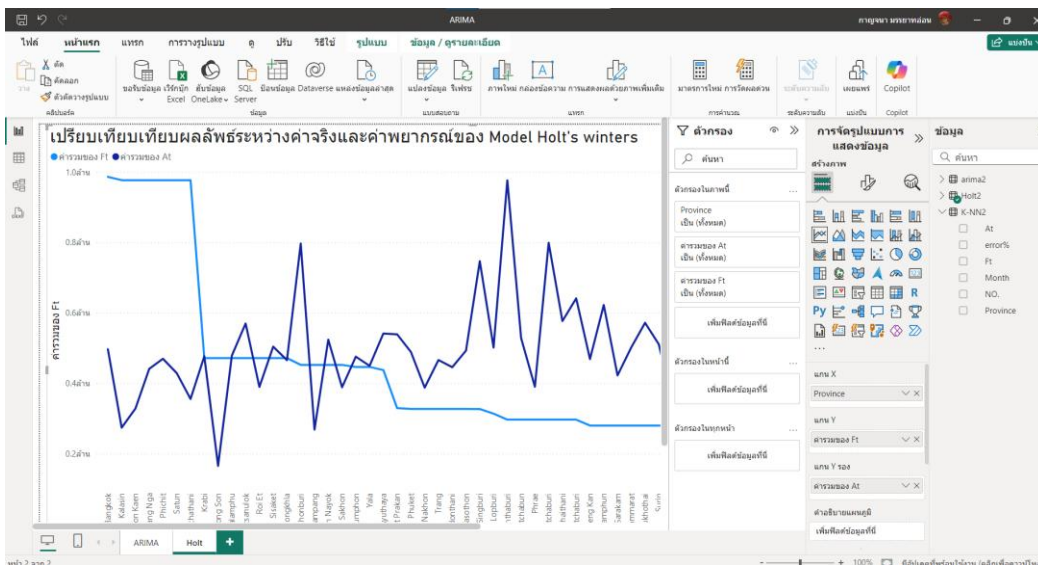
ภาพที่ 3.34 แสดงการเลือกกราฟที่เหมาะสมกับข้อมูล

6.4 ลากข้อมูล คอลัมน์ลงในแกน X และแกน Y ตามความเหมาะสมของข้อมูล

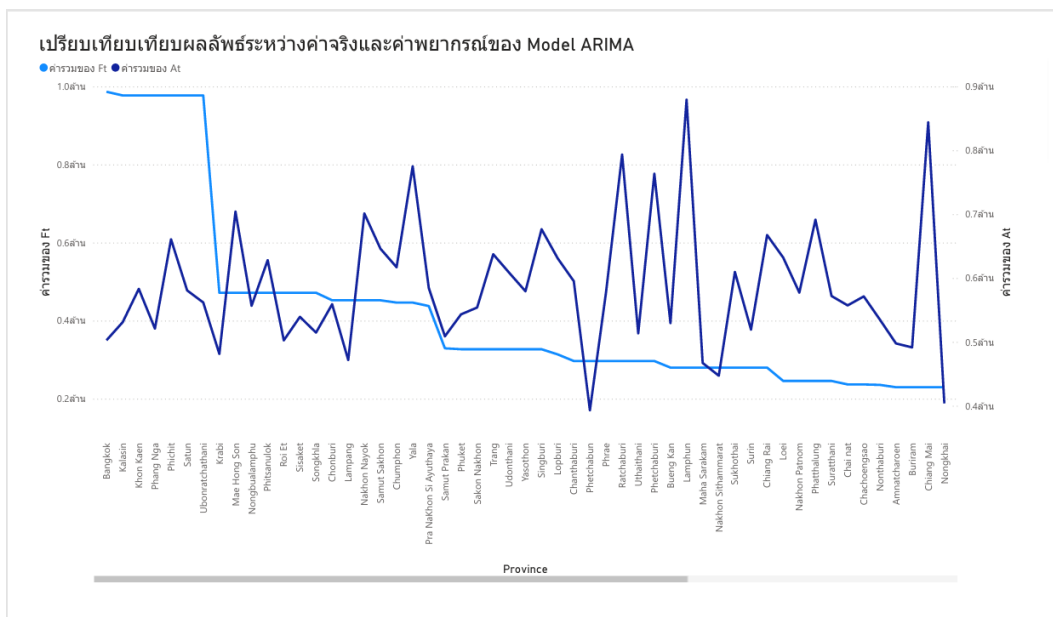


ภาพที่ 3.35 แสดงการเลือกกราฟที่เหมาะสมกับข้อมูล

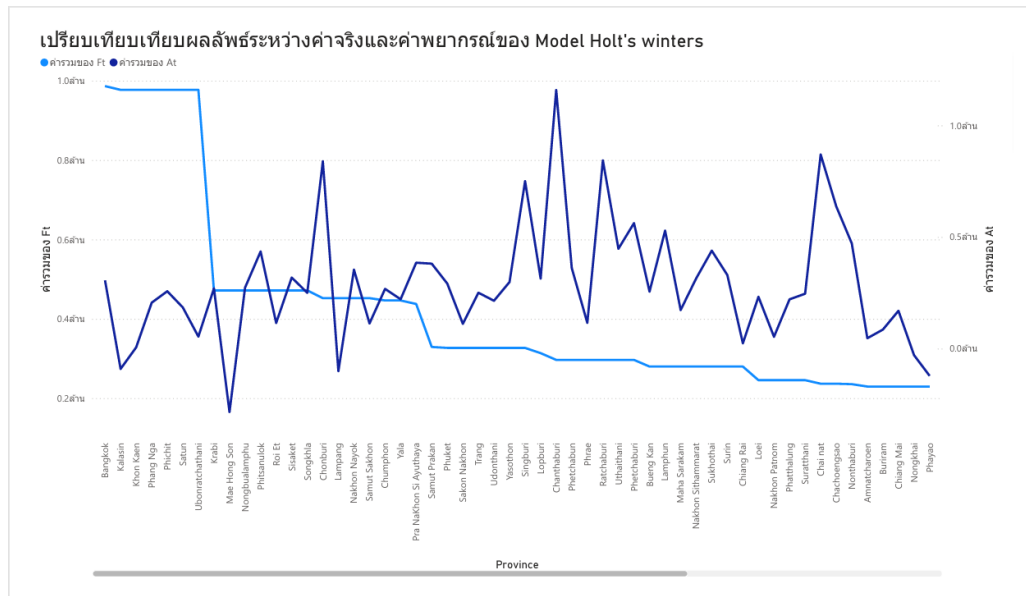
6.5 หน้าการเปรียบเทียบเทียบผลลัพธ์ระหว่างค่าจริงและค่าพยากรณ์



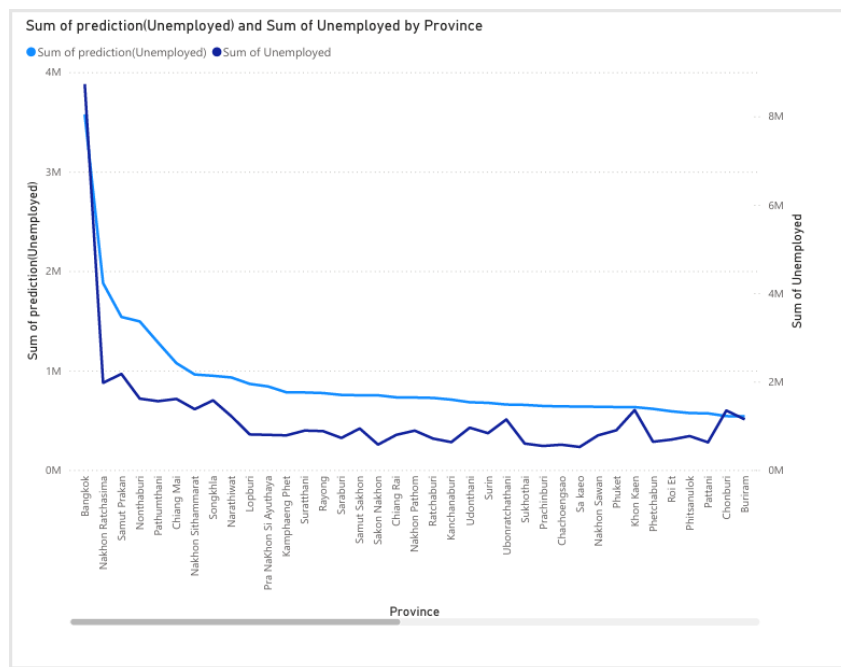
ภาพที่ 3.36 แสดงผลลัพธ์หน้าการเปรียบเทียบผลลัพธ์ระหว่างค่าจริงและค่าพยากรณ์



ภาพที่ 3.37 แสดงการเปรียบเทียบผลลัพธ์ระหว่างค่าจริงและค่าพยากรณ์ Model ARIMA



ภาพที่ 3.38 แสดงการเปรียบเทียบผลลัพธ์ระหว่างค่าจริงและค่าพยากรณ์ Model Holt's



ภาพที่ 3.39 แสดงการเปรียบเทียบผลลัพธ์ระหว่างค่าจริงและค่าพยากรณ์ Model K-NN

3.2 การออกแบบและพัฒนาเว็บไซต์

การวางแผนการจัดลำดับเนื้อหาสาระของเว็บไซต์ ออกเป็นหมวดหมู่ เพื่อจัดทำเป็นโครงสร้างในการจัดวางหน้าเว็บเพจทั้งหมด ทำให้เห็นโครงสร้างทั้งหมดของเว็บไซต์และการออกแบบโครงสร้างหรือจัดระเบียบของข้อมูลที่ชัดเจน แยกย่อยเนื้อหาออกเป็นส่วนต่าง ๆ ที่สัมพันธ์กันและให้อยู่ในมาตรฐานเดียวกัน จะช่วยให้นำมาใช้งานและง่ายต่อการเข้าอ่านเนื้อหาของผู้ใช้เว็บไซต์

3.2.1 แผนภาพบริบท (Context diagram)

3.2.1.1 ผู้ใช้ระบบ

ก) ผู้ดูแลเว็บไซต์

ข) ผู้เข้าชมเว็บไซต์

3.2.1.2 ความต้องการในเว็บไซต์

ก) ผู้ดูแลเว็บไซต์

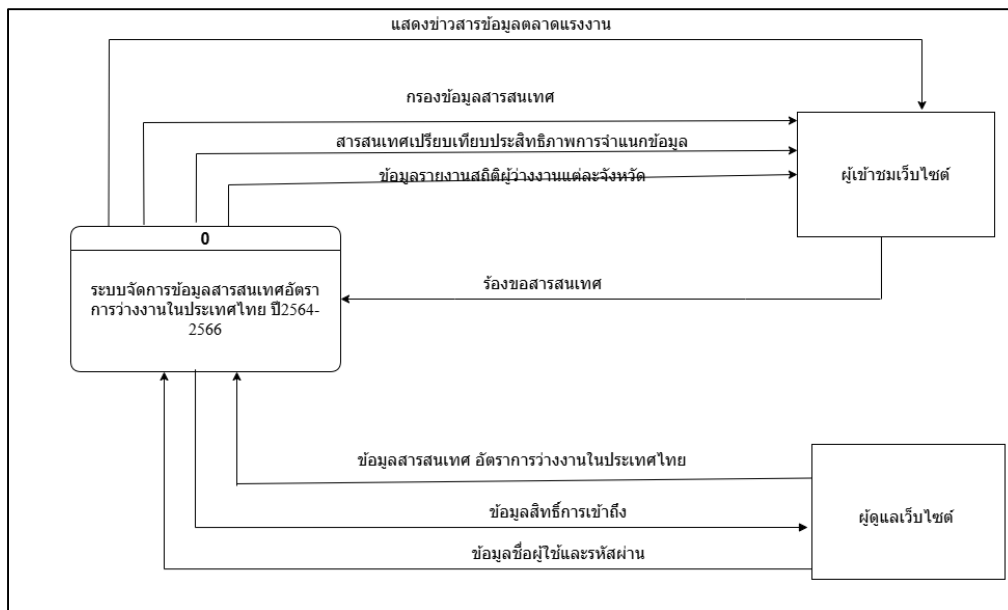
- สามารถอัปโหลดไฟล์ข้อมูลผ่านกระบวนการ CRISP-DM ลงในเว็บไซต์ และแสดงผลบนหน้าเว็บได้

ข) ผู้เข้าชมเว็บไซต์

- สามารถดูข้อมูลสารสนเทศในลักษณะตารางสรุปข้อมูลสถิติ อัตราการว่างงานในประเทศไทยในช่วงปี 2564 -2566

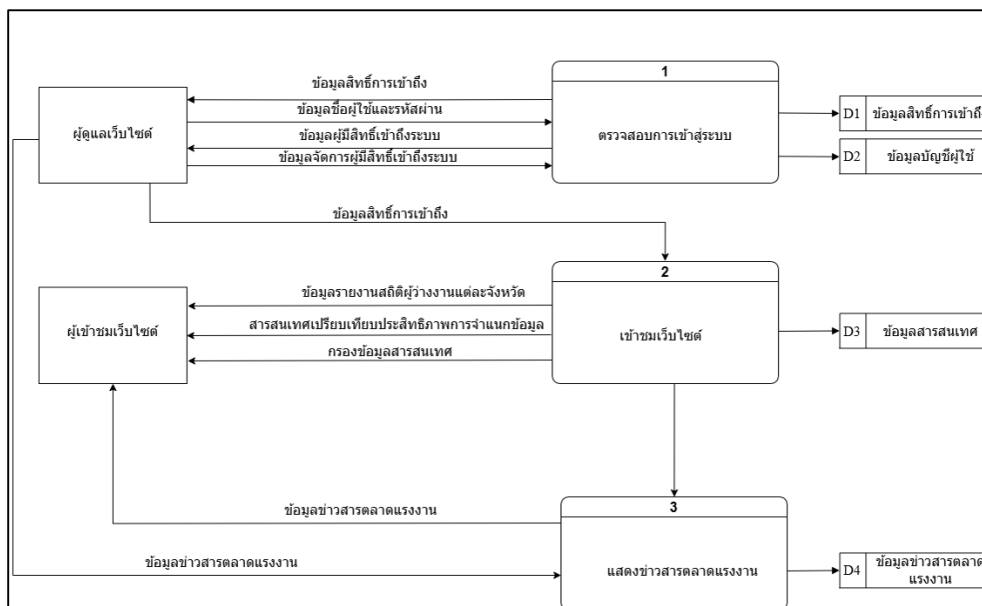
- สามารถกรองข้อมูลสารสนเทศผ่านระบบการค้นหาแสดงผล แผนภาพ สรุปข้อมูล เช่น ปี 2564-2566 และภูมิภาค

- สามารถเลือกการแสดงผลอัตราการว่างงาน แต่ละจังหวัดได้



ภาพที่ 3.40 แผนภาพบริบท (Context Diagram)

3.2.2 แผนภาพ DFD level 0 ของระบบจัดการข้อมูลสารสนเทศอัตราการทำงานในประเทศไทยปี 2564-2566



ภาพที่ 3.41 แผนภาพกระแสข้อมูลระดับที่ 0

ตารางที่ 3.3 แสดงคำอธิบายการประมวลผลของโปรเซส 1.0 : ตรวจสอบการเข้าสู่ระบบ

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศอัตรการ ว่างงานในประเทศไทย ปี2564-2566
DFD number	1.0
Process name	ตรวจสอบการเข้าสู่ระบบ
Input data flows	ข้อมูลชื่อผู้ใช้และรหัสผ่าน ข้อมูลจัดการผู้มี สิทธิ์เข้าถึงระบบ
Output data flows	ข้อมูลสิทธิ์การเข้าถึงระบบ ข้อมูลผู้มีสิทธิ์ เข้าถึงระบบ
Data stores user	ข้อมูลสิทธิ์การเข้าถึง ข้อมูลบัญชีผู้ใช้
Description	โปรเซสในการตรวจสอบสิทธิ์การเข้าใช้งานใน ระดับต่างๆ ของผู้ดูแลเว็บไซต์

ตารางที่ 3.4 แสดงคำอธิบายการประมวลผลของโปรเซส 2.0 : เข้าชมเว็บไซต์

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศอัตรการ ว่างงานในประเทศไทย ปี2564-2566
DFD number	2.0
Process name	เข้าชมเว็บไซต์
Input data flows	ข้อมูลสิทธิ์การเข้าถึง
Output data flows	ข้อมูลผู้ว่างงานแต่ละจังหวัด สารสนเทศเปรียบเทียบประสิทธิภาพการ จำแนกข้อมูล กรองข้อมูลสารสนเทศ
Data stores user	ข้อมูลสารสนเทศ
Description	โปรเซสในการเข้าชมเว็บไซต์

ตารางที่ 3.5 แสดงคำอธิบายการประมวลผลของโปรเซส 3.0 : แสดงข่าวสารตลาดแรงงาน

Process Description	
System	ระบบจัดการข้อมูลสารสนเทศอัตราการว่างงานในประเทศไทย ปี2564-2566
DFD number	3.0
Process name	แสดงข่าวสารตลาดแรงงาน
Input data flows	ข้อมูลข่าวสารตลาดแรงงาน
Output data flows	ข้อมูลข่าวสารตลาดแรงงาน
Data stores user	ข้อมูลข่าวสารตลาดแรงงาน
Description	โปรเซสแสดงข่าวสารตลาดแรงงาน

3.2.3 การออกแบบ Wireframe หน้าจอเว็บไซต์

1) หน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์



ภาพที่ 3.42 แสดงหน้าแรกของเว็บไซต์ แสดงเมนูต่าง ๆ ของหน้าเว็บไซต์

2) หน้าเว็บไซต์ที่มาจากปัญหา

การเปรียบเทียบประสิทธิภาพโมเดลเพื่อการพยากรณ์ อัตราการว่างงานในประเทศไทยในช่วงปี 2564-2566 หน้าหลัก ที่มาและปัญหา ผลการวิเคราะห์ข้อมูล ข้อมูลผู้จัดทำ

ที่มาของการวิเคราะห์ปัญหา

XX

วัตถุประสงค์

XX

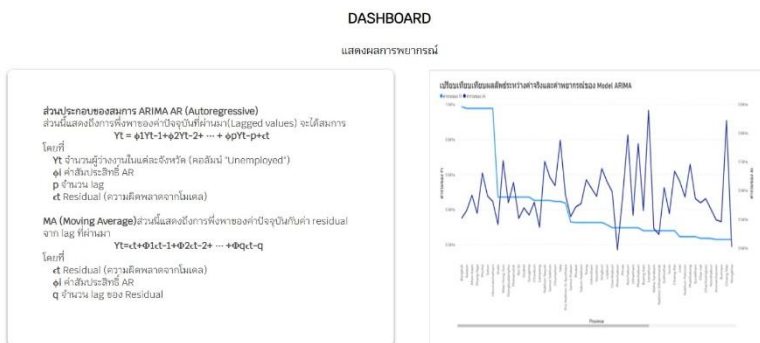
ประโยชน์ที่คาดว่าจะได้รับ

XX

ภาพที่ 3.43 แสดงหน้าเว็บไซต์ของที่มาและปัญหา

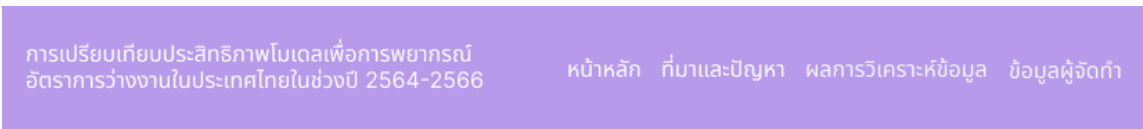
3) หน้าเว็บไซต์แสดงผลการวิเคราะห์ข้อมูล และสมการโมเดล

การเปรียบเทียบประสิทธิภาพโมเดลเพื่อการพยากรณ์ อัตราการว่างงานในประเทศไทยในช่วงปี 2564-2566 หน้าหลัก ที่มาและปัญหา ผลการวิเคราะห์ข้อมูล ข้อมูลผู้จัดทำ



ภาพที่ 3.44 แสดงหน้าเว็บไซต์แสดงผลการวิเคราะห์ข้อมูล

4)หน้าเว็บไซต์ข้อมูลผู้จัดทำ



ผู้จัดทำ



ภาพที่ 3.45 หน้าแสดงหน้าเว็บไซต์ข้อมูลผู้จัดทำ

3.3 บทสรุป

โครงการนี้วิเคราะห์แนวโน้มอัตราการว่างงานในประเทศไทยช่วงปี 2564-2566 ด้วยเทคนิคเหมืองข้อมูล (Data Mining) และกระบวนการ CRISP-DM (Cross Industry Standard Process for Data Mining) วิเคราะห์ข้อมูล 6 ขั้นตอน ความเข้าใจในธุรกิจ (Business Understanding) กำหนดเป้าหมายเพื่อวิเคราะห์ความสัมพันธ์ของข้อมูลแรงงานและว่างงานการทำความเข้าใจข้อมูล (Data Understanding) รวบรวมข้อมูลแรงงานจากเว็บไซต์กรมการจัดหางาน www.doe.go.th และตรวจสอบความครบถ้วน การเตรียมข้อมูล (Data Preparation) ทำความสะอาดข้อมูล (Data Cleaning) รวบรวมข้อมูลให้อยู่ในรูปแบบมาตรฐาน และแก้ไขข้อมูลผิดพลาด การสร้างแบบจำลอง (Modeling) ใช้โมเดล ARIMA, Holt's Winter และ k-NN ใช้ข้อมูลว่างงานเพื่อวิเคราะห์แนวโน้มและความผันผวนตามฤดูกาล เพื่อเปรียบเทียบประสิทธิภาพ การประเมินผล (Evaluation) วัดความแม่นยำของโมเดลด้วยค่า MAPE (Mean Absolute Percentage Error) และการนำไปใช้ (Deployment) นำข้อมูลและผลการวิเคราะห์ไปประยุกต์ใช้เปรียบเทียบประสิทธิภาพโมเดลเพื่อการพยากรณ์อัตราการว่างงานในประเทศไทยในช่วงปี 2564-2566