มาตรฐานอาชีพและคุณวุฒิวิชาชีพ
Occupational Standard and Professional Qualifications

สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล

จัดทำโดย สถาบันคุณวุฒิวิชาชีพ (องค์การมหาชน)

1. ชื่อมาตรฐานอาชีพ

สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล

2. ประวัติการปรับปรุงมาตรฐาน

N/A

3. ทะเบียนอ้างอิง (Imprint)

N/A

4. ข้อมูลเบื้องต้น

มาตรฐานอาชีพและคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล มุ่งเน้นเพื่อส่งเสริมและสนับสนุนให้กลุ่มสาขาอาชีพ
จัดทำมาตรฐานอาชีพและคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล ให้สอดคล้องกับความต้องการของผู้ประกอบการ
มีความเป็นสากลและเหมาะสมกับประเทศไทย เป็นที่ยอมรับทั้งภายในประเทศและระดับสากล โดยเฉพาะกลุ่มประเทศอาเซียน เพื่อรองรับประชาคมเศรษฐกิจอาเซียน
เพื่อสร้างเครือข่ายการจัดทำ พัฒนา และเผยแพร่ มาตรฐานอาชีพและคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล
และประชาสัมพันธ์ให้ระบบคุณวุฒิวิชาชีพและมาตรฐานอาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล เป็นที่รับรู้และยอมรับในทุกภาคส่วน
มาตรฐานอาชีพและคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล ฉบับนี้จัดทำโดยการวิเคราะห์ข้อมูลเกี่ยวกับสาขาวิชาชีพวิทยาศาสตร์ข้อมูล
ทั้งในประเทศและต่างประเทศ ร่วมกับวิธีเทคนิควิเคราะห์หน้าที่ (Functional Analysis) โดยจากการวิเคราะห์พบว่า มีหน้าที่หลัก (Key Function) 5 หน้าที่หลัก
ซึ่งทุกหน้าที่งานมีความสำคัญต่อการปฏิบัติงานด้านวิทยาศาสตร์ข้อมูลทั้งสิ้น ทั้งนี้ในแต่ละหน้าที่งานยังประกอบไปด้วยหลากหลายอาชีพที่เกี่ยวข้อง
อย่างไรก็ตามในมาตรฐานฉบับนี้ได้ดำเนินการจัดทำอาชีพที่มีความสำคัญเร่งด่วนเพื่อตอบสนองความต้องการของประเทศ
โดยได้ดำเนินการจัดทำมาตรฐานอาชีพและคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล ทั้งสิ้น 5 อาชีพ 13
คุณวุฒิวิชาชีพตามรายละเอียดที่อยู่ในมาตรฐานอาชีพนี้

5. ประวัติการปรับปรุงมาตรฐานในแต่ละครั้ง

N/A

6. ครั้งที่

ครั้งที่ 1

7. คุณวุฒิวิชาชีพที่ครอบคลุม (Professional Qualifications included)

สาขาวิชาชีพอุตสาหกรรมดิจิทัล

สาขาวิทยาศาสตร์ข้อมูล

อาชีพนักวิทยาศาสตร์ข้อมูล ระดับ 5

8. คุณวุฒิวิชาชีพที่เกี่ยวข้อง (Related Professional Qualifications)

N/A

9. หน่วยสมรรถนะทั้งหมดในมาตรฐานอาชีพ (List of All Units of Competence within this Occupational Standards)

รหัสหน่วยสมรรถนะ	**เนื้อหา**
70401	เลือกเทคนิควิธีสำหรับการเรียนรู้แบบมีผู้สอน
70402	เลือกเทคนิคการวิเคราะห์แบบไม่มีผู้สอน
70403	สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบมีผู้สอน
70404	สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน

10. ระดับคุณวุฒิ

10.1 สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล อาชีพนักวิทยาศาสตร์ข้อมูล ระดับ 5

คุณลักษณะของผลการเรียนรู้ (Characteristics of Outcomes)

บุคคลที่มีคุณลักษณะของผลการเรียนรู้ที่พึงประสงค์ในอาชีพ นักวิทยาศาสตร์ข้อมูล (Data Science) ระดับ 5 จะมีความรู้ด้านเศรษฐศาสตร์ กระบวนการทางสถิติ ฐานข้อมูลคอมพิวเตอร์ การวิเคราะห์ข้อมูล โปรแกรมคอมพิวเตอร์ การนำเสนอ มีทักษะในระบุประเด็นสำคัญที่มีผลต่อธุรกิจ เพื่อกำหนดความต้องการข้อมูล มีทักษะการจัดทำโครงการจัดหาข้อมูล สามารถรวบรวมข้อมูลจากแหล่งต่าง ๆทั้งในรูปแบบรายงาน และผู้มีส่วนเกี่ยวข้องสู่การสร้างชุดข้อมูลที่พร้อมใช้งาน และคัดเลือก/สร้างโมเดลการวิเคราะห์หาความเชื่อมโยงและประเด็นสำคัญจากข้อมูล เพื่อวิเคราะห์ผลและกำหนดแนวทางจัดการปัญหา รวมถึงสามารถจัดการปัญหาในบริบทที่มีการเปลี่ยนแปลงทั่วไป สามารถคิดวิเคราะห์และประเมินสถานการณ์ได้ด้วยตนเอง โดยต้องมีความรับผิดชอบต่อบทบาทความเป็นผู้นำ ให้คำแนะนำ/สอนงานและกำกับดูแลผู้ร่วมงานประเมินผลปฏิบัติงานและส่งมอบงานได้ตามเป้าหมาย
มีเทคนิคในการประยุกต์หลักการในการแก้ปัญหาทางเทคนิคหน้างานควบคู่กับการใช้คู่มือ สามารถใช้สารสนเทศเพื่อควบคุมคุณภาพงานได้ โดยต้องมีความรับผิดชอบต่องานในหน้าที่ ให้การสนับสนุนผู้ร่วมงานตัดสินใจแก้ปัญหาหน้างานและรายงานผลการทำงานอย่างต่อเนื่อง และมีจริยธรรมในการประกอบอาชีพ

การเลื่อนระดับคุณวุฒิวิชาชีพ (Qualification Pathways)

1. คุณสมบัติของผู้ที่สามารถเข้ารับการประเมินคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล นักวิทยาศาสตร์ข้อมูล (Data science) ระดับ 5
- ผู้ที่สำเร็จการศึกษาระดับปริญญาตรี และมีประสบการณ์บริหารจัดการข้อมูลอย่างน้อย 1 โครงการ หรือ
- มีประสบการณ์ทำงานเกี่ยวกับงานด้านวิทยาศาสตร์ข้อมูลไม่น้อยกว่า 2 ปี และมีประสบการณ์บริหารจัดการข้อมูลอย่างน้อย 1 โครงการ รวมถึงโครงการระหว่างศึกษา หรือ
- ได้รับการรับรองคุณวุฒิวิชาชีพ อาชีพนักวิเคราะห์ข้อมูล ระดับ 4 หรือ อาชีพผู้ปฏิบัติงานด้านวิศวกรรมข้อมูล ระดับ 4 และมีประสบการณ์บริหารจัดการข้อมูลอย่างน้อย 1 โครงการ รวมถึงโครงการระหว่างศึกษา
2. ผู้ที่จะผ่านการประเมินและได้รับการรับรองคุณวุฒิวิชาชีพ สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล นักวิทยาศาสตร์ข้อมูล (Data science) ระดับ 5
- ผ่านเกณฑ์การประเมินตามหน่วยสมรรถนะของอาชีพนักวิทยาศาสตร์ข้อมูล ระดับ 5 จำนวน 4 หน่วย

หลักเกณฑ์การต่ออายุหนังสือรับรองมาตรฐานอาชีพ

N/A

กลุ่มบุคคลในอาชีพ (Target Group)

ผู้ทำงานในกลุ่มอุตสาหกรรมเทคโนโลยีสารสนเทศและการสื่อสารและดิจิทัลคอนเทนต์ ด้านวิทยาศาสตร์ข้อมูล เช่นนักวิเคราะห์ข้อมูล นักวิเคราะห์สถิติ นักการตลาด วิศวกรข้อมูล บุคคลที่สำเร็จการศึกษาในสาขาที่เกี่ยวข้องหรือบุคคลที่สนใจในงานด้านวิทยาศาสตร์ข้อมูล

หน่วยสมรรถนะ (หน่วยสมรรถนะทั้งหมดของคุณวุฒิวิชาชีพนี้)

70401 เลือกเทคนิควิธีสำหรับการเรียนรู้แบบมีผู้สอน
70402 เลือกเทคนิคการวิเคราะห์แบบไม่มีผู้สอน
70403 สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบมีผู้สอน
70404 สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน

ตารางแผนผังแสดงหน้าที่

1. ตารางแสดงหน้าที่ 1

ประกาศใช้ ณ 20/04/2566

ตาราง 1 : FUNCTIONAL MAP แสดง KEY PURPOSE , KEY ROLES , KEY FUNCTION

ความมุ่งหมายหลัก Key Purpose	บทบาทหลัก Key Roles		หน้าที่หลัก Key Function	
คำอธิบาย	**รหัส**	**คำอธิบาย**	**รหัส**	**คำอธิบาย**
พัฒนาศักยภาพของบุคคลากรในสาขาอาชีพ ICT ให้สามารถแข่งขันและเป็นที่ยอมรับในระดับสากล	70	ปฏิบัติงานด้านวิทยาศาสตร์ข้อมูลได้ตามมาตรฐานอาชีพ	704	ตั้งสมมติฐานและพัฒนาแบบจำลอง

คำอธิบาย ตารางแผนผังแสดงหน้าที่เป็นแผนผังที่ใช้วิเคราะห์หน้าที่งานเพื่อให้ได้หน้าที่หลัก (Key Function)

2. ตารางแสดงหน้าที่ 1 (ต่อ)

ประกาศใช้ ณ 20/04/2566

ตาราง 2 : FUNCTIONAL MAP แสดง KEY FUNCTION , UNIT OF COMPETENCE , ELEMENT OF COMPETENCE

หน้าที่หลัก Key Function		หน่วยสมรรถนะ Unit of Competence		หน่วยสมรรถนะย่อย Element of Competence	
รหัส	คำอธิบาย	รหัส	คำอธิบาย	รหัส	คำอธิบาย
704	ตั้งสมมติฐานและพัฒนาแบบจำลอง	70401	เลือกเทคนิควิธีสำหรับการเรียนรู้แบบมีผู้สอน	70401.01	จำแนกเทคนิคการเรียนรู้แบบมีผู้สอนแบบต่าง ๆ
				70401.02	เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค Decision Tree
				70401.03	เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค SVM
				70401.04	เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค ANN
				70401.05	เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์การถดถอย(Regression) ด้วยเทคนิค Linear Regression
				70401.06	สื่อสารในระหว่างการดำเนินงาน
		70402	เลือกเทคนิคการวิเคราะห์แบบไม่มีผู้สอน	70402.01	จำแนกเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบต่าง ๆ
				70402.02	เลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค k-Mean
				70402.03	เลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค Hierarchical
				70402.04	เลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการค้นหากฎความสัมพันธ์ (Association) ด้วยเทคนิค Apriori
		70403	สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบมีผู้สอน	70403.01	กำหนดขั้นตอนต่าง ๆ ของแบบจำลอง
				70403.02	แบ่งข้อมูลสำหรับชุดฝึกการเรียนรู้ (Training set, Validation and Test set)
				70403.03	สร้างแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบการจำแนกข้อมูล (Classification)
				70403.04	ทดสอบแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification)
				70403.05	สร้างแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression)
				70403.06	ทดสอบแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression)
				70403.07	สรุปแบบจำลอง
		70404	สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน	70404.01	กำหนดขั้นตอนต่าง ๆ ของแบบจำลอง

หน้าที่หลัก Key Function		หน่วยสมรรถนะ Unit of Competence		หน่วยสมรรถนะย่อย Element of Competence	
รหัส	**คำอธิบาย**	**รหัส**	**คำอธิบาย**	**รหัส**	**คำอธิบาย**
704	ตั้งสมมติฐานและพัฒนาแบบจำลอง	70404	สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน	70404.02	สร้างแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการแบ่งกลุ่มข้อมูล (Clustering)
				70404.03	ทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการแบ่งกลุ่มข้อมูล (Clustering)
				70404.04	สร้างแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการค้นหากฎความสัมพันธ์ (Association)
				70404.05	ทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการค้นหากฎความสัมพันธ์ (Association)
				70404.06	สรุปผลแบบจำลอง

คำอธิบาย

ตารางแผนผังแสดงหน้าที่ (ต่อ) เป็นแผนผังที่ใช้วิเคราะห์หน้าที่งานหลังจากได้หน้าที่หลัก (Key Function) เพื่อให้ได้ หน่วยสมรรถนะ (Unit of Competence) และหน่วยสมรรถนะย่อย (Element of Competence)

1. รหัสหน่วยสมรรถนะ 70401

2. ชื่อหน่วยสมรรถนะ เลือกเทคนิควิธีสำหรับการเรียนรู้แบบมีผู้สอน

3. ทบทวนครั้งที่ 1 / 2566

4. สร้างใหม่ ☐ **ปรับปรุง** ☑

5. สำหรับชื่ออาชีพและรหัสอาชีพ (Occupational Classification)

นักวิทยาศาสตร์ข้อมูล (Data Scientist)

6. คำอธิบายหน่วยสมรรถนะ (Description of Unit of Competency)

ผู้ที่ผ่านสมรรถนะนี้จะมีความรู้เกี่ยวกับการเลือกเทคนิควิธีสำหรับการเรียนรู้แบบมีผู้สอน
โดยสามารถเลือกเทคนิคสำหรับการสร้างแบบจำลองแบบมีผู้สอนสำหรับการจำแนกข้อมูล (Classification) และการวิเคราะห์ถดถอย (Regression)
ที่มีความเหมาะสมกับเครื่องมือที่มี เพื่อใช้ประมวลผลกับข้อมูลทั้งที่มีโครงสร้างที่ชัดเจน และไม่มีโครงสร้างที่ชัดเจน ที่ซึ่งมีความสัมพันธ์กับเป้าหมายธุรกิจ

7. สำหรับระดับคุณวุฒิ

1	2	3	4	5	6	7	8
☐	☐	☐	☐	☑	☐	☐	☐

8. กลุ่มอาชีพ (Sector)

สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล (Data Science)

9. ชื่ออาชีพและรหัสอาชีพอื่นที่หน่วยสมรรถนะนี้สามารถใช้ได้ (ถ้ามี)

N/A

10. ข้อกำหนดหรือกฎระเบียบที่เกี่ยวข้อง (Licensing or Regulation Related) (ถ้ามี)

N/A

11. สมรรถนะย่อยและเกณฑ์การปฏิบัติงาน (Elements and Performance Criteria)

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70401.01 จำแนกเทคนิคการเรียนรู้แบบมีผู้สอนแบบต่าง ๆ	1. ระบุความแตกต่างของเทคนิคการวิเคราะห์แบบมีผู้สอนแบบต่าง ๆ ได้ 2. ระบุลักษณะข้อมูลที่เหมาะสมกับการเรียนรู้แบบมีผู้สอนได้ 3. ระบุความแตกต่างของคำตอบหรือผลลัพธ์ที่จะได้ระหว่างเทคนิคต่าง ๆ สำหรับการเรียนรู้แบบมีผู้สอนได้	ข้อสอบข้อเขียน
70401.02 เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค Decision Tree	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค Decision Tree ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือต้นแบบจำลองด้วยเทคนิค Decision Tree ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค Decision Tree ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70401.03 เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค SVM	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค SVM ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือแบบจำลองด้วยเทคนิค SVM ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค SVM ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน
70401.04 เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค ANN	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค ANN ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือแบบจำลองด้วยเทคนิค ANN ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค ANN ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน
70401.05 เลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์การถดถอย(Regression) ด้วยเทคนิค Linear Regression	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์การถดถอย (Regression) ด้วยเทคนิค Linear Regression ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือแบบจำลองด้วยเทคนิค Linear Regression ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์การถดถอย (Regression) ด้วยเทคนิค Linear Regression ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน
70401.06 สื่อสารในระหว่างการดำเนินงาน	1. รับและส่งสารในระหว่างการดำเนินงานได้ 2. สอบถามและตอบข้อมูลในระหว่างการดำเนินงาน ได้ 3. สามารถสื่อสารผ่านสื่อสารสนเทศช่องทางต่าง ๆได้	การสัมภาษณ์ ข้อสอบข้อเขียน

12. ความรู้และทักษะก่อนหน้าที่จำเป็น (Pre-requisite Skill & Knowledge)

N/A

13. ทักษะและความรู้ที่ต้องการ (Required Skills and Knowledge)

(ก) ความต้องการด้านทักษะ

1. สามารถใช้วิธีการวิเคราะห์และวิธีทางสถิติเพื่อช่วยเตรียมข้อมูลและเลือกข้อมูลได้
2. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผลข้อมูลขนาดใหญ่ได้
3. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผล SQL หรือ NoSQL หรือที่เกี่ยวข้องได้

(ข) ความต้องการด้านความรู้

1.เข้าใจธุรกิจและวิเคราะห์ข้อมูลให้สอดคล้องกับวัตถุประสงค์ธุรกิจ

14. หลักฐานที่ต้องการ (Evidence Guide)

(ก) หลักฐานการปฏิบัติงาน (Performance Evidence)

1. ใบรับรองการปฏิบัติงานจากสถานประกอบการ

(ข) หลักฐานความรู้ (Knowledge Evidence)

1. ใบรับรองการเข้ารับการฝึกอบรม

2. ใบประกาศนียบัตรวุฒิการศึกษา

(ค) คำแนะนำในการประเมิน

1. ผู้ประเมินตรวจประเมินเกี่ยวกับการเลือกเทคนิควิธีสำหรับการเรียนรู้แบบมีผู้สอน โดยพิจารณาจากร่องรอยหลักฐานที่เกี่ยวข้องทั้งหลักฐานการปฏิบัติงาน และหลักฐานความรู้

(ง) วิธีการประเมิน

1. พิจารณาตามหลักฐานการปฏิบัติงาน

2. พิจารณาตามหลักฐานความรู้

15. ขอบเขต (Range Statement)

(ก) คำแนะนำ

หน่วยสมรรถนะนี้เป็นการเลือกเทคนิคต่าง ๆ ของการสร้างแบบจำลองแบบมีผู้สอนอย่างสอดคล้องกับปัญหาขององค์กรและข้อมูล รวมทั้งเครื่องมือช่วยต่าง ๆ ได้อย่างเหมาะสม

(ข) คำอธิบายรายละเอียด

1.การเข้าใจเทคนิคการวิเคราะห์แบบมีผู้สอน (Supervised Learning)

Supervised Learning หรือการเรียนรู้แบบมีผู้สอน เป็นศาสตร์แขนงหนึ่งใน AI หรือปัญญาประดิษฐ์ ภายใต้หัวข้อ Machine Learning คือการทำให้คอมพิวเตอร์สามารถหาคำตอบของปัญหาได้เอง ภายหลังจากเรียนรู้จากชุดข้อมูลตัวอย่างไปแล้วระยะหนึ่ง เป็นการเน้นการเรียนรู้จากข้อมูลในอดีต เพื่อสร้างเป็นแบบจำลองหรือแบบจำลองเพื่อทำนาย หรือคาดการณ์สิ่งที่จะเกิดในอนาคต โดยแบบจำลองหรือแบบจำลองอาจเป็นสมการคณิตศาสตร์ หรือกฎต่าง ๆ โดยแบ่งเป็นกลุ่มใหญ่ ๆ ดังนี้

1) การจำแนกข้อมูล (Classification)

2) การวิเคราะห์การถดถอย (Regression)

2. การจำแนกข้อมูล (Classification) คือกระบวนการสร้างแบบจำลองจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้จากกลุ่มตัวอย่างข้อมูลที่เรียกว่าข้อมูลสอนระบบ (Training data) ที่แต่ละแถวของข้อมูลประกอบด้วยฟิลด์หรือแอทริบิวท์จำนวนมาก แอทริบิวท์นี้อาจเป็นค่าต่อเนื่อง (Continuous) หรือค่ากลุ่ม (Categorical) โดยจะมีแอทริบิวท์แบ่ง (Classifying attribute) ซึ่งเป็นตัวบ่งชี้คลาสของข้อมูล จุดประสงค์ของการจำแนกประเภทข้อมูลคือการสร้างแบบจำลองการแยกแอทริบิวท์หนึ่งโดยขึ้นกับแอทริบิวท์อื่น แบบจำลองที่ได้จากการจำแนกประเภทข้อมูลจะทำให้สามารถพิจารณาคลาสในข้อมูลที่ยังมิได้แบ่งกลุ่มในอนาคตได้ ตัวอย่างเทคนิคการจำแนกข้อมูลพื้นฐานมีดังนี้

1) Decision Tree คือแบบจำลองทางคณิตศาสตร์ที่ใช้ทำนายประเภท โดยพิจารณาจากลักษณะของวัตถุ เป็นกระบวนการสร้างต้นไม้ขึ้นเพื่อใช้ในการตัดสินใจจากข้อมูลที่มีหมวดหมู่ ต้นไม้ตัดสินใจจะประกอบไปด้วยโหนดต่าง ๆ

2) การเข้าใจเทคนิคการวิเคราะห์ SVM

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) เป็นตัวแบบที่ใช้ในการแยกแยะข้อมูล โดยจะแบ่งข้อมูลด้วยระนาบหลายมิติ ซึ่งมีความใกล้เคียงกับ ANN หลักการของ SVM คือการให้อินพุท (Input) ที่ใช้ฝึก (Train) เป็นเวคเตอร์ในสเปซ N มิติ เช่นถ้าในกรณีของ 2 มิติ และ 3 มิติ จะเป็นจุดที่อยู่ในระนาบ xy และสเปซ xyz ตามลำดับ จากนั้นทำการสร้างไฮเปอร์เพลน (Hyperplane) เพื่อแยกกลุ่มของเวคเตอร์อินพุตออกเป็นประเภทต่าง ๆ ในกรณีที่เป็น 2 มิติ และ 3 มิติ ไฮเปอร์เพลน คือเส้นตรงและระนาบตามลำดับ

3) การเข้าใจเทคนิคการวิเคราะห์ ANN (MLP)

Artificial Neural Networks (ANN) นับเป็นสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence, AI) มี Artificial Neurons ที่สามารถเรียนรู้ได้โดยใช้หลักการเช่นเดียวกับสมองมนุษย์ เรียนรู้โดยการปรับค่า Connection Weight เพื่อให้ได้ผลลัพธ์ที่มีความคลาดเคลื่อน (Error) น้อยที่สุด

3. การวิเคราะห์ถดถอย (Regression)

การทำนายค่าเชิงตัวเลขโดยการใช้หลักการทางสถิติที่เรียกว่า การถดถอย (regression) ซึ่ง Regression Analysis ถือเป็นเครื่องมือทางสถิติที่มีการประยุกต์ใช้ในการประมวลผลข้อมูลในงานวิจัยค่อนข้างมาก การวิเคราะห์การถดถอยจะเป็นแบบจำลองสำหรับการหาความสัมพันธ์ระหว่าง “predictor variable” ซึ่งเป็นค่าในแอทริบิวต่าง ๆที่ใช้ในการอธิบายคุณลักษณะของข้อมูลเรคคอร์ดหนึ่งๆ กับ “response variable”

เทคนิคการวิเคราะห์ที่จัดอยู่ในกลุ่ม Regression มีหลายเทคนิค โดยมีเทคนิคที่สำคัญคือ

1) Linear Regression

2) non-Linear Regression

4. เครื่องมือช่วยการวิเคราะห์ และการทำเหมืองข้อมูล (Data mining) มีหลากหลาย โดยมีโปรแกรมที่สำคัญดังนี้

1) โปรแกรม Weka, RapidMiner

2) ภาษา R, Python, Java

16. หน่วยสมรรถนะร่วม (ถ้ามี)

N/A

17. อุตสาหกรรมร่วม/กลุ่มอาชีพร่วม (ถ้ามี)

N/A

18. รายละเอียดกระบวนการและวิธีการประเมิน (Assessment Description and Procedure)

18.1 เครื่องมือประเมินการจำแนกเทคนิคการเรียนรู้แบบมีผู้สอนแบบต่าง ๆตามข้อกำหนดมาตรฐาน
1. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.2 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค Decision Tree ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.3 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค SVM ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.4 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ด้วยเทคนิค ANN ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.5 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์การถดถอย (Regression) ด้วยเทคนิค Linear Regression ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.6 เครื่องมือประเมินการสื่อสารในระหว่างการดำเนินงานตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน

1. รหัสหน่วยสมรรถนะ 70402

2. ชื่อหน่วยสมรรถนะ เลือกเทคนิคการวิเคราะห์แบบไม่มีผู้สอน

3. ทบทวนครั้งที่ 1 / 566

4. สร้างใหม่ ☐ **ปรับปรุง** ☑

5. สำหรับชื่ออาชีพและรหัสอาชีพ (Occupational Classification)

นักวิทยาศาสตร์ข้อมูล (Data Scientist)

6. คำอธิบายหน่วยสมรรถนะ (Description of Unit of Competency)

สามารถเลือกเทคนิคสำหรับการวิเคราะห์แบบไม่มีผู้สอนสำหรับการแบ่งกลุ่มข้อมูล (Clustering) และการค้นหากฎความสัมพันธ์ (Association) ที่มีความเหมาะสมกับเครื่องมือที่มี สำหรับข้อมูลทั้งที่มีโครงสร้างที่ชัดเจน และไม่มีโครงสร้างที่ชัดเจน ที่ซึ่งมีความสัมพันธ์กับเป้าหมายธุรกิจ

7. สำหรับระดับคุณวุฒิ

1	2	3	4	5	6	7	8
☐	☐	☐	☐	☑	☐	☐	☐

8. กลุ่มอาชีพ (Sector)

สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล (Data Science)

9. ชื่ออาชีพและรหัสอาชีพอื่นที่หน่วยสมรรถนะนี้สามารถใช้ได้ (ถ้ามี)

N/A

10. ข้อกำหนดหรือกฎระเบียบที่เกี่ยวข้อง (Licensing or Regulation Related) (ถ้ามี)

N/A

11. สมรรถนะย่อยและเกณฑ์การปฏิบัติงาน (Elements and Performance Criteria)

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70402.01 จำแนกเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบต่าง ๆ	1. ระบุเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบต่าง ๆ ได้ 2. ระบุถึงแหล่งที่มาของข้อมูลที่เหมาะสมกับเทคนิคการวิเคราะห์แบบไม่มีผู้สอนได้ 3. บอกความแตกต่างของคำตอบหรือผลลัพธ์ที่จะได้ระหว่างเทคนิคต่าง ๆ ที่จัดอยู่ในกลุ่มนี้ได้	ข้อสอบข้อเขียน
70402.02 เลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค k-Mean	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค k-Mean ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือแบบจำลองด้วยเทคนิค k-Mean ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค k-Mean ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70402.03 เลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค Hierarchical	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค Hierarchical ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือแบบจำลองด้วยด้วยเทคนิค Hierarchical ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค Hierarchical ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน
70402.04 เลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการค้นหากฎความสัมพันธ์ (Association) ด้วยเทคนิค Apriori	1. ระบุหลักการของเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการค้นหากฎความสัมพันธ์ (Association) ด้วยเทคนิค Apriori ได้ 2. บอกถึงผลลัพธ์ที่จะได้รับจากการสร้างสมการหรือแบบจำลองด้วยด้วยเทคนิค Apriori ได้ 3. กำหนดเครื่องมือช่วยในการวิเคราะห์แบบมีไม่มีผู้สอนแบบการค้นหากฎความสัมพันธ์ (Association) ด้วยเทคนิค Apriori ได้ 4. ตั้งสมมติฐานที่คาดว่าจะได้รับได้	การสัมภาษณ์ ข้อสอบข้อเขียน

12. ความรู้และทักษะก่อนหน้าที่จำเป็น (Pre-requisite Skill & Knowledge)

N/A

13. ทักษะและความรู้ที่ต้องการ (Required Skills and Knowledge)

(ก) ความต้องการด้านทักษะ

1. สามารถใช้วิธีการวิเคราะห์และวิธีทางสถิติเพื่อช่วยเตรียมข้อมูลและเลือกข้อมูลได้
2. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผลข้อมูลขนาดใหญ่ได้
3. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผล SQL หรือ NoSQL หรือที่เกี่ยวข้องได้

(ข) ความต้องการด้านความรู้

1.เข้าใจธุรกิจและวิเคราะห์ข้อมูลให้สอดคล้องกับวัตถุประสงค์ธุรกิจ

14. หลักฐานที่ต้องการ (Evidence Guide)

(ก) หลักฐานการปฏิบัติงาน (Performance Evidence)

1. ใบรับรองการปฏิบัติงานจากสถานประกอบการ

(ข) หลักฐานความรู้ (Knowledge Evidence)

1. ใบรับรองการเข้ารับการฝึกอบรม
2. ใบประกาศนียบัตรวุฒิการศึกษา

(ค) คำแนะนำในการประเมิน

1. ผู้ประเมินตรวจประเมินเกี่ยวกับการเลือกเทคนิควิธีสำหรับการเรียนรู้แบบไม่มีผู้สอน โดยพิจารณาจากร่องรอยหลักฐานที่เกี่ยวข้องทั้งหลักฐานการปฏิบัติงานและหลักฐานความรู้

(ง) วิธีการประเมิน

1. พิจารณาตามหลักฐานการปฏิบัติงาน
2. พิจารณาตามหลักฐานความรู้

15. ขอบเขต (Range Statement)

(ก) คำแนะนำ

หน่วยสมรรถนะนี้เป็นการเลือกเทคนิคต่าง ๆ ของการสร้างแบบจำลองแบบไม่มีผู้สอนอย่างสอดคล้องกับปัญหาขององค์กรและข้อมูล รวมทั้งเครื่องมือช่วยต่าง ๆ ได้อย่างเหมาะสม

(ข) คำอธิบายรายละเอียด

1. การเข้าใจเทคนิคการวิเคราะห์แบบไม่มีผู้สอน (Unsupervised Learning) คือการเรียนรู้แบบระบบจะไม่ได้รับการป้อนรูปแบบผลลัพธ์ที่ต้องการไปในการเรียนรู้ และต้องเรียนรู้จากข้อมูลที่ได้จากการกระทำต่าง ๆ เอง และพัฒนาดีขึ้นเมื่อเรียนรู้จากความผิดพลาดในอดีตบ่อยครั้งขึ้น โดยสามารถแบ่งแยกย่อยได้เป็นดังนี้

1) การจัดกลุ่มข้อมูล (Clustering)

2) การค้นหาความสัมพันธ์ (Association Rule)

2. การจัดกลุ่มข้อมูล (Clustering) เป็นเทคนิคที่ใช้จำแนกหรือแบ่ง Case ของข้อมูล หรือแบ่งตัวแปรออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป Case ที่อยู่ในกลุ่มเดียวกันจะมีคุณลักษณะที่เหมือนกันหรือคล้ายกัน ส่วน Case ที่อยู่ต่างกลุ่มกัน จะมีลักษณะที่แตกต่างกัน โดยสามารถแบ่งกลุ่มหรือจัดกลุ่มได้ตามความเหมือน (Similarity) หรือตามความแตกต่าง (Dissimilarity or Distance) ของข้อมูลนั้น ๆ การจัดกลุ่มจะแตกต่างจากการแบ่งประเภทข้อมูล (Classification) โดยจะแบ่งกลุ่มข้อมูลจากความคล้าย โดยไม่มีการกำหนดคลาส (Class) ประเภทข้อมูลไว้ก่อนหรือไม่ทราบจำนวนกลุ่มล่วงหน้า โดยมีเทคนิคที่สำคัญดังนี้

1) k-Mean หรือเรียกอีกอย่างหนึ่งว่า การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) หรือ การแบ่งส่วน (Partitioning) เป็นอัลกอริทึมเทคนิคการเรียนรู้โดยไม่มีผู้สอนที่ง่ายที่สุด

2) Hierarchical หรือการแบ่งกลุ่มแบบลำดับระดับ เป็นการแบ่งกลุ่มจากกลุ่มย่อยที่ถูกแบ่งไว้ก่อนหน้านั้นซ้ำหลายครั้ง เป็นเทคนิคที่นิยมใช้กันมากในการจัดกลุ่ม Case หรือจัดกลุ่มตัวแปร

3. การค้นหาความสัมพันธ์ (Association Rule) เป็นเทคนิคหนึ่งของ Data Mining เพื่อการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำไปหารูปแบบที่เกิดขึ้นบ่อยๆ (frequent pattern) และใช้ในการวิเคราะห์ความสัมพันธ์หรือทำนายปรากฏการณ์ต่าง ๆ โดยมีเทคนิคที่สำคัญดังนี้

1) Apriorism เป็นอัลกอริทึมพื้นฐานที่ใช้ในการหาความสัมพันธ์ของข้อมูลโดยใช้หลักการค้นหาแบบวงกว้างก่อนนับทรานแซคชัน

4.เครื่องมือช่วยการวิเคราะห์ และการทำเหมืองข้อมูล (Data mining) มีหลากหลาย โดยมีโปรแกรมที่สำคัญดังนี้

1) โปรแกรม Weka, RapidMiner

2) ภาษา R, Python, Java

16. หน่วยสมรรถนะร่วม (ถ้ามี)

N/A

17. อุตสาหกรรมร่วม/กลุ่มอาชีพร่วม (ถ้ามี)

N/A

18. รายละเอียดกระบวนการและวิธีการประเมิน (Assessment Description and Procedure)

18.1 เครื่องมือประเมินการจำแนกเทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบต่าง ๆ ตามข้อกำหนดมาตรฐาน

1. ผลข้อสอบข้อเขียน

ดูรายละเอียดจากคู่มือประเมิน

18.2 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค k-Mean ตามข้อกำหนดมาตรฐาน

1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน

ดูรายละเอียดจากคู่มือประเมิน

18.3 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการแบ่งกลุ่มข้อมูล (Clustering) ด้วยเทคนิค Hierarchical ตามข้อกำหนดมาตรฐาน

1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน

ดูรายละเอียดจากคู่มือประเมิน

18.4 เครื่องมือประเมินการเลือกใช้เทคนิคการวิเคราะห์แบบไม่มีผู้สอนแบบการค้นหากฎความสัมพันธ์ (Association) ด้วยเทคนิค Apriori ตามข้อกำหนดมาตรฐาน

1. แบบฟอร์มประเมินผลการสัมภาษณ์
2. ผลข้อสอบข้อเขียน

ดูรายละเอียดจากคู่มือประเมิน

1. รหัสหน่วยสมรรถนะ 70403

2. ชื่อหน่วยสมรรถนะ สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบมีผู้สอน

3. ทบทวนครั้งที่ 1 / 2566

4. สร้างใหม่ ☐ **ปรับปรุง** ☑

5. สำหรับชื่ออาชีพและรหัสอาชีพ (Occupational Classification)

นักวิทยาศาสตร์ข้อมูล (Data Scientist)

6. คำอธิบายหน่วยสมรรถนะ (Description of Unit of Competency)

สามารถวางแผน กำหนดขั้นตอนการทดสอบแบบจำลอง การแบ่งข้อมูลสำหรับชุดฝึกการเรียนรู้ (Training set) และชุดทดสอบ (Test set) รวมทั้ง Validation set หรือวิธีอื่นที่เหมาะสมกับเทคนิคที่เลือก พร้อมทั้งสร้างและทดสอบแบบจำลองด้วยเครื่องมือช่วย (Tools) สำหรับเทคนิคการวิเคราะห์แบบมีผู้สอน

7. สำหรับระดับคุณวุฒิ

1	2	3	4	5	6	7	8
☐	☐	☐	☐	☑	☐	☐	☐

8. กลุ่มอาชีพ (Sector)

สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล (Data Science)

9. ชื่ออาชีพและรหัสอาชีพอื่นที่หน่วยสมรรถนะนี้สามารถใช้ได้ (ถ้ามี)

N/A

10. ข้อกำหนดหรือกฎระเบียบที่เกี่ยวข้อง (Licensing or Regulation Related) (ถ้ามี)

N/A

11. สมรรถนะย่อยและเกณฑ์การปฏิบัติงาน (Elements and Performance Criteria)

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70403.01 กำหนดขั้นตอนต่าง ๆ ของแบบจำลอง	1. ระบุวิธีการสร้างและทดสอบแบบจำลองได้ 2. กำหนดขั้นตอน กระบวนการ ในการทดสอบแบบจำลองได้	ข้อสอบข้อเขียน
70403.02 แบ่งข้อมูลสำหรับชุดฝึกการเรียนรู้ (Training set, Validation and Test set)	1. ระบุวิธีการแบ่งข้อมูลเพื่อทำการทดสอบแบบจำลองได้ 2. อธิบายความแตกต่างของวิธีการแบ่งข้อมูลเพื่อทำการทดสอบแบบ Self Consistency Test, Split Test และ Cross-validation Test ได้ 3. กำหนดขั้นตอนการแบ่งข้อมูลเพื่อทำการทดสอบแบบจำลองโดยวิธี Self Consistency Test, Split Test และ Cross-validation Test ได้ 4. สามารถเลือกใช้วิธีการแบ่งข้อมูลเพื่อทำการทดสอบแบบจำลองกับข้อมูลได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70403.03 สร้างแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบการจำแนกข้อมูล (Classification)	1. ใช้เครื่องมือช่วย (Tools) เพื่อสร้างแบบจำลองได้ 2. สามารถกำหนดค่าตัวแปรสำหรับในเครื่องมือช่วย (Tools) สำหรับเทคนิค การจำแนกข้อมูล (Classification) ที่เลือกใช้ได้ 3. จัดทำเอกสารรายงานประกอบการกำหนดค่าตัวแปรได้ 4. สามารถดำเนินการฝึกการเรียนรู้ (Training) กับข้อมูลที่เลือกได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70403.04 ทดสอบแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification)	1. สามารถดำเนินการตรวจสอบประสิทธิภาพ (Validation) เทคนิคที่เลือกได้ 2. สามารถดำเนินการทดสอบแบบจำลอง (Testing) เทคนิคที่เลือกได้ 3. สามารถตรวจสอบสถานะแบบจำลองที่เกิดจากการปรับแก้ปัญหา Overfitting หรือ Underfitting ได้ 4. สามารถปรับค่าตัวแปรของแบบจำลอง (regularization) ได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70403.05 สร้างแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression)	1. ใช้เครื่องมือช่วย (Tools) เพื่อสร้างแบบจำลองได้ 2. สามารถกำหนดค่าตัวแปรสำหรับในเครื่องมือช่วย (Tools) สำหรับเทคนิคการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression) ที่เลือกใช้ได้ 3. จัดทำเอกสารรายงานประกอบการกำหนดค่าตัวแปรได้ 4. สามารถดำเนินการฝึกการเรียนรู้ (Training) กับข้อมูลที่เลือกได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70403.06 ทดสอบแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression)	1. สามารถดำเนินการตรวจสอบประสิทธิภาพ (Validation) เทคนิคที่เลือกได้ 2. สามารถดำเนินการทดสอบแบบจำลอง (Testing) เทคนิคที่เลือกได้ 3. สามารถตรวจสอบสถานะแบบจำลองที่เกิดจากการปรับแก้ปัญหา Overfitting หรือ Underfitting ได้ 4. สามารถปรับค่าตัวแปรของแบบจำลอง (regularization) ได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70403.07 สรุปแบบจำลอง	1. บันทึกการตั้งค่าตัวแปรสำหรับเทคนิคที่เลือกใช้ได้ 2. อธิบายคุณลักษณะเฉพาะของแบบจำลองเพื่อใช้ต่อไปได้ 3. สามารถสรุปผลการใช้แบบจำลองกับข้อมูลและผลลัพธ์ที่ได้	ข้อสอบข้อเขียน

12. ความรู้และทักษะก่อนหน้าที่จำเป็น (Pre-requisite Skill & Knowledge)

N/A

13. ทักษะและความรู้ที่ต้องการ (Required Skills and Knowledge)

(ก) ความต้องการด้านทักษะ

1. สามารถใช้วิธีการวิเคราะห์และวิธีทางสถิติเพื่อช่วยเตรียมข้อมูลและเลือกข้อมูลได้
2. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผลข้อมูลขนาดใหญ่ได้
3. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผล SQL หรือ NoSQL หรือที่เกี่ยวข้องได้

(ข) ความต้องการด้านความรู้

1. เข้าใจธุรกิจและวิเคราะห์ข้อมูลให้สอดคล้องกับวัตถุประสงค์ธุรกิจ

14. หลักฐานที่ต้องการ (Evidence Guide)

(ก) หลักฐานการปฏิบัติงาน (Performance Evidence)

1. ใบรับรองการปฏิบัติงานจากสถานประกอบการ

(ข) หลักฐานความรู้ (Knowledge Evidence)

1. ใบรับรองการเข้ารับการฝึกอบรม

2. ใบประกาศนียบัตรวุฒิการศึกษา

(ค) คำแนะนำในการประเมิน

1. ผู้ประเมินตรวจประเมินเกี่ยวกับการสร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบมีผู้สอน โดยพิจารณาจากร่องรอยหลักฐานที่เกี่ยวข้องทั้งหลักฐานการปฏิบัติงานและหลักฐานความรู้

(ง) วิธีการประเมิน

1. พิจารณาตามหลักฐานการปฏิบัติงาน

2. พิจารณาตามหลักฐานความรู้

15. ขอบเขต (Range Statement)

(ก) คำแนะนำ

หน่วยสมรรถนะนี้เป็นการเลือกและออกแบบเพื่อการแบ่งชุดข้อมูล ให้เหมาะสมกับเทคนิคแบบจำลองที่เลือกและจำนวนข้อมูลตัวอย่างที่มีสำหรับการเรียนรู้แบบมีผู้สอน

(ข) คำอธิบายรายละเอียด

1.เข้าใจและบอกข้อดี ข้อเสียของการแบ่งข้อมูลแบบต่าง ๆ เพื่อนำทดสอบประสิทธิภาพของแบบจำลอง

ในการทดสอบประสิทธิภาพของแบบจำลอง เพื่อให้ทราบถึงความน่าเชื่อถือ การวัดประสิทธิภาพ ก่อนนำแบบจำลองไปใช้ทางธุรกิจถือว่ามีความจำเป็น โดยทั่วไปเป็นการแบ่งข้อมูลออกเป็นชุดเพื่อการทดสอบ

ทั้งนี้การแบ่งข้อมูลเพื่อทำการทดสอบมี 3 วิธีที่นิยม ดังนี้

1) วิธี Self Consistency หรือบางครั้งเรียกว่า Use Training Set นี้เป็นวิธีการที่ง่ายที่สุด

2) วิธี Split Test การแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยข้อมูลส่วนที่หนึ่ง (70% หรือ 80%) ใช้ในการสร้างแบบจำลอง และข้อมูลส่วนที่สอง (30% หรือ 20%) ใช้ใน การทดสอบประสิทธิภาพของแบบจำลอง

3) วิธี Cross-validation Test เป็นวิธีที่นิยมในการทำงานวิจัย เพื่อใช้ในการทดสอบประสิทธิภาพของแบบจำลองเนื่องจากผลที่ได้มีความน่าเชื่อถือ

2. การวัด error ของ model คือการทดสอบประสิทธิภาพของแบบจำลองก่อนการทดสอบ

3. การสอน model ด้วย Training set โดยใช้ Training set

4. การเข้าใจเครื่องมือช่วย (Tools) สำหรับทดสอบแบบจำลอง โดยสามารถบอกวิธีใช้เครื่องมือช่วย รวมทั้งตั้งค่าเริ่มต้นได้

5. การใช้เครื่องมือช่วย สำหรับการ โดยสามารถใช้เครื่องมือช่วย รวมทั้งตั้งค่าเริ่มต้น และทดสอบเครื่องมือช่วยกับแบบจำลองที่เลือกใช้กับข้อมูลที่มีและได้รับการจัดการให้เหมาะสมกับแบบจำลองแล้ว โดยโปรแกรมเครื่องมือช่วยมีดังนี้

1) โปรแกรม Weka, RapidMiner

2) ภาษา R, Python, Java

16. หน่วยสมรรถนะร่วม (ถ้ามี)

N/A

17. อุตสาหกรรมร่วม/กลุ่มอาชีพร่วม (ถ้ามี)

N/A

18. รายละเอียดกระบวนการและวิธีการประเมิน (Assessment Description and Procedure)

18.1 เครื่องมือประเมินการกำหนดขั้นตอนต่าง ๆ ของแบบจำลองตามข้อกำหนดมาตรฐาน
1. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.2 เครื่องมือประเมินการแบ่งข้อมูลสำหรับชุดฝึกการเรียนรู้ (Training set, Validation and Test set) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.3 เครื่องมือประเมินการสร้างแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบการจำแนกข้อมูล (Classification) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.4 เครื่องมือประเมินการทดสอบแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการจำแนกข้อมูล (Classification) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.5 เครื่องมือประเมินการสร้างแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.6 เครื่องมือประเมินการทดสอบแบบจำลองโดยใช้เครื่องมือ (Tools) สำหรับการวิเคราะห์แบบมีผู้สอนแบบการวิเคราะห์ถดถอย (Regression) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.7 เครื่องมือประเมินการสรุปผลแบบจำลองตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน

1. รหัสหน่วยสมรรถนะ 70404

2. ชื่อหน่วยสมรรถนะ สร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน

3. ทบทวนครั้งที่ 1 / 2566

4. สร้างใหม่ ☐ **ปรับปรุง** ☑

5. สำหรับชื่ออาชีพและรหัสอาชีพ (Occupational Classification)

นักวิทยาศาสตร์ข้อมูล (Data Scientist)

6. คำอธิบายหน่วยสมรรถนะ (Description of Unit of Competency)

ผู้ที่ผ่านสมรรถนะนี้จะมีความรู้เกี่ยวกับแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน โดยสามารถวางแผน กำหนดขั้นตอนการทดสอบแบบจำลอง การแบ่งข้อมูลสำหรับชุดฝึกการเรียนรู้ (Training set) และชุดทดสอบ (Test set) หรือวิธีอื่นที่เหมาะสมกับเทคนิคที่เลือก พร้อมทั้งสร้างและทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน

7. สำหรับระดับคุณวุฒิ

1	2	3	4	5	6	7	8
☐	☐	☐	☐	☑	☐	☐	☐

8. กลุ่มอาชีพ (Sector)

สาขาวิชาชีพอุตสาหกรรมดิจิทัล สาขาวิทยาศาสตร์ข้อมูล (Data Science)

9. ชื่ออาชีพและรหัสอาชีพอื่นที่หน่วยสมรรถนะนี้สามารถใช้ได้ (ถ้ามี)

N/A

10. ข้อกำหนดหรือกฎระเบียบที่เกี่ยวข้อง (Licensing or Regulation Related) (ถ้ามี)

N/A

11. สมรรถนะย่อยและเกณฑ์การปฏิบัติงาน (Elements and Performance Criteria)

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70404.01 กำหนดขั้นตอนต่าง ๆ ของแบบจำลอง	1. ระบุวิธีการสร้างและทดสอบแบบจำลองได้ 2. กำหนดขั้นตอน กระบวนการ ในการทดสอบแบบจำลองได้	ข้อสอบข้อเขียน
70404.02 สร้างแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการแบ่งกลุ่มข้อมูล (Clustering)	1. ใช้เครื่องมือ (Tools) เพื่อสร้างและทดสอบแบบจำลองได้ 2. สามารถกำหนดค่าตัวแปรสำหรับในเครื่องมือ (Tools) สำหรับเทคนิคการแบ่งกลุ่มข้อมูล (Clustering) ที่เลือกใช้ได้ 3. จัดทำเอกสารรายงานประกอบการกำหนดค่าตัวแปรได้ 4. สามารถดำเนินการสร้างแบบจำลองการแบ่งกลุ่มข้อมูลได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70404.03 ทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการแบ่งกลุ่มข้อมูล (Clustering)	1. สามารถตรวจสอบประสิทธิภาพแบบจำลอง ได้ 2. สามารถปรับค่าตัวแปรของแบบจำลองได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70404.04 สร้างแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการค้นหากฎความสัมพันธ์ (Association)	1. ใช้เครื่องมือ (Tools) เพื่อสร้างและทดสอบแบบจำลองได้ 2. สามารถกำหนดค่าตัวแปรสำหรับในเครื่องมือ (Tools) สำหรับเทคนิคการค้นหากฎความสัมพันธ์ (Association) ที่เลือกใช้ได้ 3. จัดทำเอกสารรายงานประกอบการกำหนดค่าตัวแปรได้ 4. สามารถดำเนินการสร้างการค้นหากฎความสัมพันธ์ข้อมูลได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน
70404.05 ทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการค้นหากฎความสัมพันธ์ (Association)	1. สามารถตรวจสอบประสิทธิภาพแบบจำลองได้ 2. สามารถปรับค่าตัวแปรของแบบจำลองได้	การสาธิตการปฏิบัติงาน ข้อสอบข้อเขียน

สมรรถนะย่อย (Element)	เกณฑ์ในการปฏิบัติงาน (Performance Criteria)	วิธีการประเมิน (Assessment)
70404.06 สรุปผลแบบจำลอง	1. บันทึกการตั้งค่าตัวแปรสำหรับเทคนิคที่เลือกใช้ได้ 2. อธิบายและรายงานคุณลักษณะเฉพาะของแบบจำลองเพื่อใช้ต่อไปได้ 3. อธิบายพฤติกรรมของแบบจำลองที่เลือกใช้ได้ 4. สามารถวิเคราะห์ผลการใช้แบบจำลองกับข้อมูลและผลลัพธ์ที่ได้	ข้อสอบข้อเขียน

12. ความรู้และทักษะก่อนหน้าที่จำเป็น (Pre-requisite Skill & Knowledge)

N/A

13. ทักษะและความรู้ที่ต้องการ (Required Skills and Knowledge)

(ก) ความต้องการด้านทักษะ

1. สามารถใช้วิธีการวิเคราะห์และวิธีทางสถิติเพื่อช่วยเตรียมข้อมูลและเลือกข้อมูลได้
2. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผลข้อมูลขนาดใหญ่ได้
3. สามารถใช้เทคนิคและเครื่องมือสำหรับประมวลผล SQL หรือ NoSQL หรือที่เกี่ยวข้องได้

(ข) ความต้องการด้านความรู้

1.เข้าใจธุรกิจและวิเคราะห์ข้อมูลให้สอดคล้องกับวัตถุประสงค์ธุรกิจ

14. หลักฐานที่ต้องการ (Evidence Guide)

(ก) หลักฐานการปฏิบัติงาน (Performance Evidence)

1. ใบรับรองการปฏิบัติงานจากสถานประกอบการ

(ข) หลักฐานความรู้ (Knowledge Evidence)

1. ใบรับรองการเข้ารับการฝึกอบรม
2. ใบประกาศนียบัตรวุฒิการศึกษา

(ค) คำแนะนำในการประเมิน

1. ผู้ประเมินตรวจประเมินเกี่ยวกับการสร้างแบบจำลองสำหรับเทคนิคการวิเคราะห์แบบไม่มีผู้สอน โดยพิจารณาจากร่องรอยหลักฐานที่เกี่ยวข้องทั้งหลักฐานการปฏิบัติงานและหลักฐานความรู้

(ง) วิธีการประเมิน

1. พิจารณาตามหลักฐานการปฏิบัติงาน
2. พิจารณาตามหลักฐานความรู้

15. ขอบเขต (Range Statement)

(ก) คำแนะนำ

หน่วยสมรรถนะนี้เป็นการเลือกและออกแบบเพื่อการแบ่งชุดข้อมูล ให้เหมาะสมกับเทคนิคแบบจำลองที่เลือกและจำนวนข้อมูลตัวอย่างที่มีสำหรับการเรียนรู้แบบมีผู้สอน

(ข) คำอธิบายรายละเอียด

1. การสร้างแบบจำลองแบบไม่มีผู้สอน และการวัด error ของ model คือการทดสอบประสิทธิภาพของแบบจำลองก่อนการ
2. การเข้าใจเครื่องมือช่วย (Tools) สำหรับทดสอบแบบจำลอง โดยสามารถบอกวิธีใช้เครื่องมือช่วย รวมทั้งตั้งค่าเริ่มต้นได้
3. การใช้เครื่องมือช่วย สำหรับการ โดยสามารถใช้เครื่องมือช่วย รวมทั้งตั้งค่าเริ่มต้น และทดสอบเครื่องมือช่วยกับแบบจำลองที่เลือกใช้กับข้อมูลที่มีและได้รับการจัดการให้เหมาะสมกับแบบจำลองแล้ว โดยโปรแกรมเครื่องมือช่วยมีดังนี้

1) โปรแกรม Weka, RapidMiner

2) ภาษา R, Python, Java

16. หน่วยสมรรถนะร่วม (ถ้ามี)

N/A

17. อุตสาหกรรมร่วม/กลุ่มอาชีพร่วม (ถ้ามี)

N/A

18. รายละเอียดกระบวนการและวิธีการประเมิน (Assessment Description and Procedure)

18.1 เครื่องมือประเมินการกำหนดขั้นตอนต่าง ๆ ของแบบจำลองตามข้อกำหนดมาตรฐาน
1. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.2 เครื่องมือประเมินการสร้างแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการแบ่งกลุ่มข้อมูล (Clustering) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.3 เครื่องมือประเมินการทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการแบ่งกลุ่มข้อมูล (Clustering) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.4 เครื่องมือประเมินการสร้างแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการค้นหากฎความสัมพันธ์ (Association) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.5 เครื่องมือประเมินการทดสอบแบบจำลองด้วยเครื่องมือ (Tools) สำหรับการค้นหากฎความสัมพันธ์ (Association) ตามข้อกำหนดมาตรฐาน
1. แบบฟอร์มประเมินผลการสาธิตการปฏิบัติงาน
2. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน
18.6 เครื่องมือประเมินการสรุปผลแบบจำลองตามข้อกำหนดมาตรฐาน
1. ผลข้อสอบข้อเขียน
ดูรายละเอียดจากคู่มือประเมิน