

6411011660027

สิรภัทร วงศ์พัฒน์เสวก

ให้นักศึกษา Review Paper ที่เกี่ยวข้องกับ Classification พร้อมตอบคำถาม
ดังต่อไปนี้

ชื่อ Paper : การเปรียบเทียบเทคนิคการจำแนกประเภทข้อมูลสำหรับทำนายแนวโน้มความสำเร็จ
การศึกษาของนักเรียน

ผู้แต่ง : ศรินทรา เสือพิทักษ์ , บุญมี นิสสัยดี และ วิภาวรรณ บัวทอง

สำนักพิมพ์/ชื่อวารสาร และปีที่ตีพิมพ์ :

วารสารวิชาการชาชนันท์ มรภ.ภูเก็ต ปีที่ 5 ฉบับที่ 2 พ.ศ. 2564

PKRU SciTech Journal 5(2) 2021

จุดประสงค์หลักของ Paper นี้ ผู้เขียนต้องการทำอะไร :

- เพื่อสร้างแบบจำลองในการทำนายแนวโน้มความสำเร็จการศึกษาระดับประกาศนียบัตรวิชาชีพ
ของวิทยาลัยอาชีวศึกษาภูเก็ต ด้วยเทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล

ชื่ออัลกอริธึม Classification ที่ใช้ :

- อัลกอริทึมนาอีฟเบย์ (naive bayes)
- อัลกอริทึมต้นไม้ตัดสินใจ (decision tree)
- อัลกอริทึมเพื่อนบ้านใกล้เคียง (K – nearest neighbor : KNN)

อธิบายทฤษฎี (เฉพาะอัลกอริธึม Classification ที่ใช้) :

- อัลกอริทึมนาอีฟเบย์ (naive bayes) เป็นเครื่องมือที่อาศัยหลักการของความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes theorem) จำแนกข้อมูลโดยการเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่
- อัลกอริทึมต้นไม้ตัดสินใจ (decision tree) เป็นอัลกอริทึมที่ใช้สร้างต้นไม้ตัดสินใจ พัฒนาโดย Ross Quinlan โดย จะสร้างผังต้นไม้ตัดสินใจมาจากกลุ่มของข้อมูลชุดฝึกที่ใช้หลักการของทฤษฎีสารสนเทศ (information entropy) ในการเลือกโหนดโดยใช้ค่าความถูกต้องของคุณลักษณะแต่ละตัวเพื่อใช้แยกเส้นทางตัดสินใจ โดยพิจารณาจากค่าทฤษฎีสารสนเทศ
- อัลกอริทึมเพื่อนบ้านใกล้เคียง (K – nearest neighbor : KNN) เป็นวิธีการที่ได้รับความนิยมในการใช้งานอย่างมาก เนื่องจากเป็นวิธีที่ง่ายและมีประสิทธิภาพซึ่งสามารถนำไปประยุกต์ใช้กับงานได้หลากหลาย เช่น งานทางด้านการจำแนกกลุ่ม (classification)

Attribute หรือ Field ที่ใช้ (อธิบาย) :

ลำดับ	แอตทริบิวต์	ชนิดข้อมูล	รายละเอียด
1	no	ตัวเลข	ลำดับ
2	Id_student	ตัวเลข	รหัสนักเรียน
3	Sex	ตัวอักษร	เพศ
4	Thai	ตัวเลข	เกรดวิชาภาษาไทยพื้นฐาน
5	Language	ตัวเลข	เกรดวิชาภาษาอังกฤษในชีวิตจริง
6	Science	ตัวเลข	เกรดวิชาวิทยาศาสตร์เพื่อพัฒนาทักษะชีวิต
7	Math	ตัวเลข	เกรดวิชาคณิตศาสตร์พื้นฐานอาชีพ
8	Social	ตัวเลข	เกรดวิชาประวัติศาสตร์ชาติไทย
9	Gymnastics	ตัวเลข	เกรดวิชาพลศึกษาเพื่อพัฒนาสุขภาพ
10	Computer	ตัวเลข	เกรดวิชาคอมพิวเตอร์และสารสนเทศเพื่องานอาชีพ
11	Account	ตัวเลข	เกรดวิชาการบัญชีเบื้องต้น
12	Marketing	ตัวเลข	เกรดวิชาการขายเบื้องต้น
13	Typewriter	ตัวเลข	เกรดวิชาพิมพ์ไทยเบื้องต้น
14	GPA1	ตัวเลข	คะแนนเฉลี่ยเทอม 1
15	GPA2	ตัวเลข	คะแนนเฉลี่ยเทอม 2
16	GPA_finish	ตัวเลข	คะแนนเฉลี่ย
17	Status	ตัวอักษร	สถานะ

ผลลัพธ์ที่ได้ (อาจเป็นรูปหรือตารางที่ผู้วิจัยใช้อธิบาย) :

ตารางที่ 3 เปรียบเทียบผลการวิเคราะห์ข้อมูลตามการจำแนกประเภทข้อมูลด้วยเทคนิคเหมืองข้อมูล

อัลกอริทึม	เกณฑ์ในการวัดค่าประสิทธิภาพ	วิธีการประเมินประสิทธิภาพของแบบจำลอง		
		การตรวจสอบแบบข้ามจำนวน 5 กลุ่ม (ร้อยละ)	การตรวจสอบแบบข้ามจำนวน 10 กลุ่ม (ร้อยละ)	การสุ่มด้วยการแบ่งเป็นร้อยละ 66 ที่ร้อยละ 66 (ร้อยละ)
นาอ์ฟเบย์	ค่าความแม่นยำ	86.30	86.30	81.40
	ค่าความถูกต้อง	86.00	86.00	81.20
	ค่าเรียกคืน	86.30	86.30	81.40
	ค่าประสิทธิภาพโดยรวม	85.60	85.60	79.70
ต้นไม้ตัดสินใจ	ค่าความแม่นยำ	87.40	87.40	84.70
	ค่าความถูกต้อง	87.50	87.50	87.40
	ค่าเรียกคืน	87.40	87.40	84.70
	ค่าประสิทธิภาพโดยรวม	86.70	86.70	82.70
เพื่อนบ้าน	ค่าความแม่นยำ	87.40	87.40	79.70
ใกล้เคียง	ค่าความถูกต้อง	87.80	87.80	80.80
(IBK = 3)	ค่าเรียกคืน	87.40	87.40	79.70
	ค่าประสิทธิภาพโดยรวม	86.50	86.50	76.50

อธิบายผลลัพธ์ (ตามที่ผู้วิจัยเขียนไว้) :

งานวิจัยนี้ได้ใช้เทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล ผลการเปรียบเทียบประสิทธิภาพทั้ง 3 อัลกอริทึม พบว่าอัลกอริทึมต้นไม้ตัดสินใจ (decision tree) ซึ่งสร้างแบบจำลองจากการตรวจสอบแบบข้าม จำนวน 5 กลุ่ม และการตรวจสอบแบบข้าม จำนวน 10 กลุ่ม ให้ค่าประสิทธิภาพของแบบจำลองได้ดีที่สุด คือ ค่าความแม่นยำ ร้อยละ 87.40 ค่าความถูกต้อง ร้อยละ 87.50 ค่าเรียกคืน ร้อยละ 87.40 และค่าประสิทธิภาพโดยรวม ร้อยละ 86.70 ซึ่งสามารถนำมาใช้เป็นแบบจำลองเพื่อทำนายแนวโน้มการสำเร็จการศึกษาของนักเรียนได้