มหาวิทยาลัยสุโขทัยธรรมาธิราช

กิจกรรมประจำชุดวิชา
99204 สถิติสำหรับวิทยาการข้อมูล
ภาคการศึกษา 1/2568

สาขาวิชาวิทยาศาสตร์และเทคโนโลยี

คำนำ

เนื่องด้วยมหาวิทยาลัยสุโขทัยธรรมาธิราช มุ่งให้ผู้เรียนและนักศึกษาได้มีส่วนร่วมในกระบวนการศึกษาเล่าเรียนครบวงจร ตั้งแต่ก่อนเรียน ระหว่างเรียน และหลังจากเรียนเสร็จสิ้นไปแล้ว โดยจัดระบบการประเมินครบทั้ง 3 ส่วน ทั้งการประเมินก่อนเรียน ระหว่างเรียน และประเมินผลสุดท้าย

การประเมินกิจกรรม เป็นส่วนหนึ่งของการวัดผลสัมฤทธิ์ทางการเรียนสุดท้าย จึงให้ผู้เรียนและนักศึกษาทำกิจกรรมภาคปฏิบัติตามที่กำหนดให้โดยมีวัตถุประสงค์เพื่อให้ผู้เรียนและนักศึกษามีความสามารถ ดังนี้

1. สรุปหรือประมวลเนื้อหาสาระของเอกสารการสอนทั้งชุดวิชาหรือกลุ่มเนื้อหากลุ่มใดกลุ่มหนึ่ง
2. ประยุกต์ความรู้จากเอกสารการสอนเพื่อจัดทำโครงการพัฒนางานอย่างใดอย่างหนึ่งที่นักศึกษาทำ
3. พัฒนาระบบ โครงการ ชิ้นงาน ฯลฯ ตามกระบวนการหรือขั้นตอนที่แสดงไว้ในหน่วยใดหน่วยหนึ่งของเอกสารการสอน
4. คิด วิเคราะห์ นำเสนอข้อมูลและความคิดในเชิงสร้างสรรค์

นอกจากนี้การทำกิจกรรมประจำชุดวิชายังทำให้นักศึกษาได้ศึกษาเอกสารการสอนตั้งแต่ต้นภาคการศึกษา และจากการวิจัยพบว่านักศึกษาที่ทำกิจกรรมจะมีโอกาสสอบผ่านในปลายภาคมากกว่านักศึกษาที่ไม่ทำกิจกรรม

คณะกรรมการบริหารชุดวิชาสถิติสำหรับวิทยากรข้อมูลขอให้นักศึกษาทุกท่านประสบความสำเร็จในการศึกษาชุดวิชานี้ และสามารถนำความรู้ไปเป็นประโยชน์ต่อการดำเนินชีวิต และการทำงานสืบไป

คณะกรรมการบริหาร
99204 สถิติสำหรับวิทยากาข้อมูล

1. การประเมินผล

เกณฑ์การให้คะแนนกิจกรรมพิจารณาจากการตอบที่ตรงประเด็นคำถาม การครอบคลุมความถูกต้องของคำตอบ ความชัดเจนของการนำเสนอ และความละเอียดประณีตของชิ้นงาน

มหาวิทยาลัยไม่บังคับให้นักศึกษาทุกคนต้องทำกิจกรรม นักศึกษาอาจเลือกทำหรือไม่ทำก็ได้ โดยการประเมินปลายภาคสำหรับชุดวิชานี้ แบ่งออกเป็น 2 กรณี

กรณีที่ 1 นักศึกษาทำกิจกรรม ในกรณีนี้มหาวิทยาลัยแบ่งคะแนนออกเป็น 2 ส่วน ส่วนแรกจากคะแนนสอบปลายภาคคิดร้อยละ 80 และส่วนที่สองจากคะแนนกิจกรรมคิดร้อยละ 20 โดยคะแนนกิจกรรมจะนำไปใช้ในการประเมินทั้งการสอบไล่และสอบซ่อม นักศึกษาที่มิได้ส่งกิจกรรมในการสอบไล่ไม่สามารถส่งกิจกรรมเพื่อเป็นคะแนนในการสอบซ่อม

กรณีที่ 2 นักศึกษาไม่ทำกิจกรรม ในกรณีนี้มหาวิทยาลัยประเมินผลจากการสอบปลายภาคเพียงอย่างเดียว

ในการประเมินผลปลายภาค นักศึกษากลุ่มที่ทำกิจกรรมและไม่ทำกิจกรรมได้รับประเมินโดยใช้ข้อสอบฉบับเดียวกัน นักศึกษากลุ่มที่ทำกิจกรรมมีคะแนนเต็ม 80 คะแนน ส่วนนักศึกษากลุ่มที่ไม่ทำกิจกรรมจะมีคะแนนเต็ม 100 คะแนน สำหรับนักศึกษาที่ทำกิจกรรมเพื่อให้นักศึกษาได้ประโยชน์สูงสุด มหาวิทยาลัยจะนำคะแนนสอบปลายภาคของนักศึกษาเพียงอย่างเดียวมาเปรียบเทียบกับการคิดคะแนนสอบปลายภาครวมกับคะแนนกิจกรรม แล้วนำคะแนนส่วนที่มากกว่าไปใช้ในการตัดสินผลการสอบให้แก่นักศึกษา ดังตัวอย่างต่อไปนี้

<u>ตัวอย่างที่ 1</u> นักศึกษาได้คะแนนกิจกรรม 18 คะแนน และทำข้อสอบได้ 70 ข้อ (คิดเป็น $\frac{70}{120}\times 80$ เท่ากับ 46.67 คะแนน) นักศึกษาจะได้คะแนนกิจกรรมรวมกับคะแนนสอบปลายภาค 18 + 46.67 เท่ากับ 64.67 คะแนน กรณีคิดคะแนนจากการสอบปลายภาคเพียงอย่างเดียว นักศึกษาจะได้ $70\times\frac{100}{120}$ เท่ากับ 58.33 คะแนน ดังนั้นมหาวิทยาลัยจะเลือกให้นักศึกษาได้คะแนน 64.67 คะแนน

<u>ตัวอย่างที่ 2</u> นักศึกษาได้คะแนนกิจกรรม 13 คะแนน และทำข้อสอบได้ 92 ข้อ (คิดเป็น $\frac{92}{120}\times 80$ เท่ากับ 61.33 คะแนน) นักศึกษาจะได้คะแนนกิจกรรมรวมกับคะแนนสอบปลายภาค 13 + 61.33 เท่ากับ 74.33 คะแนน กรณีคิดคะแนนจากการสอบปลายภาคเพียงอย่างเดียว นักศึกษาจะได้ $92\times\frac{100}{120}$ เท่ากับ 76.67 คะแนน ดังนั้นมหาวิทยาลัยจะเลือกให้นักศึกษาได้คะแนน 76.67 คะแนน

2. การส่งกิจกรรมประจำชุด

ให้นักศึกษาดำเนินการ ดังนี้

1. ให้นักศึกษาส่ง**กิจกรรมประจำชุดวิชาไปยังมหาวิทยาลัยผ่านระบบออนไลน์** และสำเนาไฟล์กิจกรรมที่ทำเสร็จแล้วไว้เป็นหลักฐาน ทางเว็บไซต์ https://eduapp.stou.ac.th/practice
2. ให้นักศึกษา **แสดงวิธีการหาคำตอบ** ของโจทย์ในกิจกรรม โดย **เขียนด้วยลายมือตนเอง เท่านั้น ให้นักศึกษาทำกิจกรรมด้วยตนเอง *ถ้าตรวจสอบได้ว่ามีการลอกกัน หรือไม่ได้ใช้ความรู้ของตนเอง หรือวิธีทำเหมือนกันหลายฉบับ จะไม่ตรวจให้คะแนน***

   การเขียนด้วยลายมือตัวเองนั้นนักศึกษาอาจเขียนลงกระดาษแล้วถ่ายรูป หรือใช้ปากกาดิจิทัลเขียนบน pad และ save เป็นไฟล์รูปภาพ หรือ ไฟล์ pdf
3. กำหนดส่งกิจกรรมประจำชุดวิชา **ภายในวันที่ 31 ตุลาคม 2568**
4. นักศึกษาสามารถตรวจสอบว่าสำนักบริการการศึกษาได้รับกิจกรรมที่นักศึกษาส่งไปแล้วหรือยัง โดยโทรศัพท์สอบถามได้ที่หมายเลข 02-982-9633 หรือโทรศัพท์ติดต่อสำนักบริการการศึกษา หมายเลข 02-504-7621 หรือโทรศัพท์ติดต่อศูนย์สารสนเทศ หมายเลข 02-504-7788 มือถือ 084-360-4465, 084-439-9478, 084-360-5612 และ 084-360-4957 หรือที่ E-mail : ic.proffice@stou.ac.th

3. เนื้อหากิจกรรม

<u>คำชี้แจง</u> ให้นักศึกษา **แสดงวิธีการหาคำตอบ** ของโจทย์ในกิจกรรมต่อไปนี้ โดย **เขียนด้วยลายมือตนเองเท่านั้น ให้นักศึกษาทำกิจกรรมด้วยตนเอง *<u>ถ้าตรวจสอบได้ว่ามีการลอกกัน หรือไม่ได้ใช้ความรู้ของตนเอง หรือวิธีทำเหมือนกันหลายฉบับ จะไม่ตรวจให้คะแนน</u>*** โดยตอบในกระดาษ A4 เท่านั้น โจทย์ในกิจกรรมมีทั้งหมด **15 ข้อ**

ข้อ 1. ความรู้เบื้องต้นเกี่ยวกับสถิติและวิทยาการข้อมูล

จงบรรยายความสัมพันธ์ระหว่างวิทยาการข้อมูล (Data Science) กับศาสตร์ด้านอื่น ๆ โดยให้ครอบคลุมประเด็นต่าง ๆ ต่อไปนี้

- ☐ การประยุกต์ใช้วิทยาการข้อมูลในศาสตร์ต่าง ๆ
- ☐ วิธีที่วิทยาการข้อมูลช่วยเพิ่มประสิทธิภาพในการทำงานของศาสตร์นั้น ๆ
- ☐ การบูรณาการระหว่างวิทยาการข้อมูลกับศาสตร์อื่น ๆ
- ☐ ตัวอย่างการใช้วิทยาการข้อมูลในแต่ละศาสตร์

ข้อ 2. เครื่องมือสำหรับวิเคราะห์ข้อมูลสถิติ

จงเปรียบเทียบภาษา R และ Python โดยให้พิจารณาประเด็นด้านต่าง ๆ ต่อไปนี้

- ☐ จุดเด่นและจุดด้อยของทั้งสองภาษา
- ☐ การใช้งานในบริบทที่ต่างกัน
- ☐ การสนับสนุนและชุมชนผู้ใช้
- ☐ ความยืดหยุ่นและความสะดวกในการเรียนรู้
- ☐ ตัวอย่างการใช้งานที่เป็นที่นิยมของทั้งสองภาษา

ข้อ 3. สถิติเชิงพรรณนาและการวิเคราะห์สำรวจข้อมูล

จงบรรยายเกี่ยวกับสถิติเชิงพรรณนา (Descriptive Statistics) และการวิเคราะห์สำรวจข้อมูล (Exploratory Data Analysis: EDA) โดยให้พิจารณาด้านต่าง ๆ ต่อไปนี้:

- ☐ ความหมายและความสำคัญของสถิติเชิงพรรณนาและการวิเคราะห์สำรวจข้อมูล
- ☐ เทคนิคและเครื่องมือที่ใช้ในการวิเคราะห์
- ☐ ขั้นตอนการดำเนินการในแต่ละกระบวนการ
- ☐ ตัวอย่างการใช้งานและผลลัพธ์ที่ได้
- ☐ ความแตกต่างและความสัมพันธ์ระหว่างสถิติเชิงพรรณนาและการวิเคราะห์สำรวจข้อมูล

ข้อ 4. ความน่าจะเป็นและตัวแปรสุ่ม

จงบรรยายเกี่ยวกับตัวแปรสุ่มแบบต่อเนื่อง (Continuous Random Variables) และตัวแปรสุ่มแบบไม่ต่อเนื่อง (Discrete Random Variables) โดยให้ครอบคลุมประเด็นด้านต่าง ๆ ต่อไปนี้

- ☐ ความหมายและลักษณะของตัวแปรสุ่มแบบต่อเนื่องและไม่ต่อเนื่อง
- ☐ ความแตกต่างระหว่างตัวแปรสุ่มแบบต่อเนื่องและไม่ต่อเนื่อง

ข้อ 5. การแจกแจงเชิงสถิติ

จากการสำรวจความคิดเห็นของประชาชนพบว่า โอกาสที่คน ๆ หนึ่งจะสนับสนุนมาตรการเกี่ยวกับการควบคุมคุณภาพสินค้าครั้งใหม่มีเท่ากับ 0.6 ($p = 0.6$) ถ้าสำรวจความคิดเห็นของประชาชนจำนวน 25 คน ($n = 25$) จงแสดงการคำนวณความน่าจะเป็นที่จะมีคนสนับสนุนมาตรการใหม่น้อยกว่า 15 คน

ข้อ 6. การสุ่มตัวอย่าง

จงอธิบายวิธีการคำนวณขนาดตัวอย่างที่จำเป็น (Sample Size) เพื่อให้มั่นใจว่าผลลัพธ์ที่ได้มีความเชื่อถือสูงพอที่จะใช้ในการวิเคราะห์ทางสถิติ

ข้อ 7. สถิติเชิงอนุมาน

จงอธิบายความหมายของการเชิงอนุมาน (Inference) ในทางสถิติ และอธิบายการทำงานของขั้นตอนในการทำสถิติเชิงอนุมาน พร้อมทั้งยกตัวอย่างการใช้การเชิงอนุมานในทางสถิติเพื่อการตัดสินใจหรือการวิเคราะห์ข้อมูล

ข้อ 8. การวิเคราะห์ความแปรปรวน

จงอธิบายการใช้การวิเคราะห์ความแปรปรวนในการวิจัยทางสังคมศาสตร์หรือทางการเมือ พร้อมทั้งยกตัววอย่างการประยุกต์ใช้การวิเคราะห์ความแปรปรวนเพื่อให้เห็นภาพรวมของผลลัพธ์วิจัยที่ทำได้

ข้อ 9. สถิติไม่อิงพารามิเตอร์

จงอธิบายความหมายของสถิติที่ไม่อิงพารามิเตอร์ (non-parametric statistics) พร้อมทั้งยกตัวอย่างของสถิติที่ไม่อิงพารามิเตอร์ที่นิยมใช้และสามารถประยุกต์ใช้ได้ในสถานการณ์ต่าง ๆ

ข้อ 10. การวิเคราะห์สหสัมพันธ์และการถดถอยเชิงเส้น

จงคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (Pearson's correlation coefficient) ระหว่างตัวแปร X และ Y ในชุดข้อมูลต่อไปนี้ พร้อมทั้งอธิบายว่าค่าสัมประสิทธิ์สหสัมพันธ์ที่คำนวณได้หมายถึงอะไร

- ☐ X: [10, 12, 15, 18, 20]
- ☐ Y: [5, 6, 8, 9, 10]

ข้อ 11. การวิเคราะห์อนุกรมเวลาและการพยากรณ์

จงอธิบายแนวทางการพยากรณ์อนุกรมเวลา (Time Series Forecasting) และเปรียบเทียบระหว่างวิธีการพยากรณ์แบบใช้ค่าเฉลี่ยเคลื่อนที่ (Moving Average) และแบบใช้การถดถอยเชิงเส้น (Linear Regression) โดยแจกแจงข้อดีและข้อเสียของแต่ละวิธี

ข้อ 12. การจัดการข้อมูลสำหรับงานด้านสถิติและวิทยาการข้อมูล

จงอธิบายผลกระทบของข้อมูลสูญหาย (Missing Values) และค่านอกเกณฑ์ (Outliers) ต่อผลลัพธ์ของการวิเคราะห์ข้อมูล เช่น การคาดคะเน การวิเคราะห์เชิงสถิติ และการสร้างแบบจำลอง พร้อมยกตัวอย่างสถานการณ์ที่ค่า Missing Values และ Outliers ส่งผลกระทบต่อผลลัพธ์การวิเคราะห์

ข้อ 13. หลักการทางสถิติกับงานด้านการเรียนรู้ของเครื่อง

จากข้อมูลต่อไปนี้

ลักษณะ	สี	ขนาด	รูปร่าง	คลาส
1	แดง	เล็ก	กลม	A
2	แดง	ใหญ่	สี่เหลี่ยม	B
3	เขียว	เล็ก	กลม	A
4	เขียว	เล็ก	สี่เหลี่ยม	B
5	เขียว	ใหญ่	กลม	A
6	แดง	ใหญ่	กลม	A

จงอธิบายวิธีการคำนวณความน่าจะเป็นในแบบจำลอง Naive Bayes ของการที่วัตถุสีแดง ขนาดเล็ก รูปร่างกลม จะอยู่ในคลาส A

ข้อ 14. การวัดประสิทธิภาพและการแปลผลการเรียนรู้ของเครื่อง

จากข้อมูลการทดสอบแบบจำลองการจำแนกประเภท จงคำนวณและวาดเมทริกซ์ความสับสน (Confusion Matrix) สำหรับแบบจำลองที่ได้ผลการทดสอบดังนี้

- ☐ จำนวนข้อมูลทั้งหมด: 200
- ☐ จำนวนข้อมูลที่ทำนายถูกต้อง (True Positive): 120
- ☐ จำนวนข้อมูลที่ทำนายผิด (False Positive): 20
- ☐ จำนวนข้อมูลที่ทำนายถูกต้องต่อคลาสอื่น (False Negative): 30
- ☐ จำนวนข้อมูลที่ทำนายผิดต่อคลาสอื่น (True Negative): 30

ข้อ 15. การประยุกต์ใช้งานสถิติสำหรับวิทยาการข้อมูล

จงอธิบายความแตกต่างของการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Learning) และไม่มีผู้สอน (Unsupervised Learning)

ปกรายงาน

กิจกรรมประจำชุดวิชา

99204 สถิติสำหรับวิทยาการข้อมูล

ภาคการศึกษาที่ 1/2568

สาขาวิชาวิทยาศาสตร์และเทคโนโลยี

ชื่อนักศึกษา.....

รหัสประจำตัวนักศึกษา | | | | | | | | | | |

ที่อยู่ .....

.....

โทรศัพท์ (ถ้ามี) .....

ข้าพเจ้าขอยอมรับการตัดสินผลคะแนนภาคปฏิบัติจากผู้ประเมินเป็นที่สุด

ลงชื่อ.....

(.....)