**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา**

**ข้อสอบปลายภาค ภาคการศึกษาที่ 1/2568**

**รายวิชา:** ENGCE207 Advanced Topics in Computer Engineering

**เวลาสอบ:** 180 นาที

**คะแนนเต็ม:** 250 คะแนน

**คำชี้แจง:**

1. ข้อสอบมีทั้งหมด 10 ข้อ
2. โปรดเขียนคำตอบลงในกระดาษคำตอบที่จัดเตรียมไว้ให้
3. ไม่อนุญาตให้นำอุปกรณ์สื่อสารทุกชนิดเข้าห้องสอบ
4. อนุญาตให้ใช้เครื่องคิดเลขพื้นฐานได้
5. โปรดเขียนชื่อ-นามสกุล และรหัสนักศึกษาให้ชัดเจน

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**1.Linear Regression (25 คะแนน) (แสดงวิธีการคำนวณอย่างละเอียด)**

จากข้อมูลความสัมพันธ์ระหว่างชั่วโมงการติวหนังสือและคะแนนสอบของนักศึกษา 4 คน ดังตาราง:

|  |  |
| --- | --- |
| **ชั่วโมงติว (X)** | **คะแนนสอบ (Y)** |
| 2 | 65 |
| 4 | 75 |
| 5 | 80 |
| 7 | 90 |

จงคำนวณหาสมการเส้นตรง (y = mx + c) ที่แสดงความสัมพันธ์นี้ และทำนายคะแนนสอบของนักศึกษาที่ใช้เวลาติว 6 ชั่วโมง

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**2. K-Nearest Neighbors - k-NN (25 คะแนน) (แสดงวิธีการคำนวณอย่างละเอียด)**

กำหนดให้มีข้อมูล 2 คลาส (A และ B) และมีจุดข้อมูลใหม่ที่พิกัด (5, 6) หากกำหนดให้ใช้ k=3 จงคำนวณหาระยะทางแบบยูคลิด (Euclidean Distance) จากจุดใหม่ไปยังจุดอื่นๆ ทั้งหมด และระบุว่าจุดข้อมูลใหม่ควรถูกจัดอยู่ในคลาสใด

* P1: (1, 5) - คลาส A
* P2: (2, 7) - คลาส A
* P3: (8, 4) - คลาส B
* P4: (7, 1) - คลาส B
* P5: (9, 5) - คลาส B

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**3. K-Means Clustering (25 คะแนน) (แสดงวิธีการคำนวณอย่างละเอียด)**

กำหนดให้มีข้อมูล 4 จุดคือ P1(1,1), P2(1,2), P3(6,7), P4(7,7) และมี Centroid เริ่มต้น 2 ตัวคือ C1(1,1) และ C2(6,7) จงแสดงการทำงานของอัลกอริทึม K-Means จนกระทั่ง Centroid หยุดการเปลี่ยนแปลง (Convergence) และสรุปผลลัพธ์สุดท้ายต่อไปนี้:

3.1. การจัดกลุ่มในแต่ละรอบ (Assignment Step)

3.2. ตำแหน่ง Centroid ใหม่ในแต่ละรอบ (Update Step)

3.3. กลุ่มของจุดข้อมูลสุดท้าย และตำแหน่งของ Centroid สุดท้าย

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**4. Model Evaluation (25 คะแนน) (แสดงวิธีการคำนวณอย่างละเอียด)**

หลังจากสร้างแบบจำลองสำหรับคัดกรองอีเมลสแปม ท่านได้ผลลัพธ์เป็น Confusion Matrix ดังนี้:

|  |  |  |
| --- | --- | --- |
|  | **ทำนายว่า: Not Spam** | **ทำนายว่า: Spam** |
| **ค่าจริง: Not Spam** | 180 (TN) | 15 (FP) |
| **ค่าจริง: Spam** | 25 (FN) | 80 (TP) |

จากตาราง Confusion Matrix ข้างต้น จงคำนวณค่าต่อไปนี้ :

4.1 Accuracy

4.2 Precision ของคลาส "Spam"

4.3 Recall ของคลาส "Spam"

4.4 F1-Score ของคลาส "Spam"

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**5. Principal Component Analysis - PCA (25 คะแนน) (แสดงวิธีการคำนวณอย่างละเอียด)**

จากชุดข้อมูลที่มี 2 มิติ (X, Y) จำนวน 4 จุดข้อมูล ดังตาราง:

|  |  |  |
| --- | --- | --- |
| **จุดข้อมูล** | **X** | **Y** |
| **A** | 1 | 2 |
| **B** | 3 | 3 |
| **C** | 4 | 5 |
| **D** | 6 | 6 |

จงแสดงขั้นตอนการทำ PCA เพื่อลดมิติข้อมูลให้เหลือ 1 มิติ:

5.1 คำนวณหาค่าเฉลี่ยและปรับข้อมูล (Center the Data)

5.2 คำนวณหาเมทริกซ์ความแปรปรวนร่วมเกี่ยว (Covariance Matrix)

5.3 จงใช้ PC1 นี้เพื่อแปลงข้อมูลที่ปรับแล้วทั้ง 4 จุดจาก 2 มิติ ไปยังมิติใหม่ (1D) **(คำใบ้: Eigenvector ที่สำคัญที่สุด (PC1) ของเมทริกซ์ในข้อ 5.2 คือ [0.79, 0.61])**

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**6. Decision Tree (25 คะแนน) (แสดงวิธีการคำนวณอย่างละเอียด)**

จากข้อมูลการอนุมัติสินเชื่อโดยพิจารณาจาก "รายได้" และ "ประวัติเครดิต" ดังตาราง จงคำนวณหาค่า Standard Deviation Reduction (SDR) เพื่อตัดสินใจว่าควรใช้เงื่อนไขใดในการแบ่งข้อมูลที่ Root Node เป็นเงื่อนไขแรก ระหว่าง: ()

6.1 รายได้ <= 40,000

6.2 ประวัติเครดิต == 'ดี'

|  |  |  |
| --- | --- | --- |
| **รายได้ (บาท)** | **ประวัติเครดิต** | **ยอดสินเชื่อที่ให้ (Y)** |
| 30,000 | ดี | 100,000 |
| 50,000 | ดี | 200,000 |
| 25,000 | ไม่ดี | 50,000 |
| 60,000 | ดี | 250,000 |
| 45,000 | ไม่ดี | 150,000 |

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**7. จริยธรรมปัญญาประดิษฐ์ (AI Ethics) (25 คะแนน)**

จากเนื้อหาในบทที่ 10 เรื่อง NLP และจริยธรรม AI จงเลือกประเด็นปัญหาทางจริยธรรมมา 1 ประเด็น (เช่น อคติใน AI, ความโปร่งใส, ความเป็นส่วนตัว) อธิบายด้วยความเข้าใจของตนเองว่าปัญหานั้นคืออะไร และยกตัวอย่างสถานการณ์ที่อาจเกิดขึ้นได้จริง พร้อมเสนอแนวทางแก้ไขเบื้องต้น

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**8. การเปรียบเทียบอัลกอริทึม (Algorithm Comparison) (25 คะแนน)**

จงเปรียบเทียบหลักการทำงานของอัลกอริทึม K-Nearest Neighbors (k-NN) และ Support Vector Machine (SVM) สำหรับปัญหาการจำแนกประเภท (Classification) พร้อมทั้งอภิปรายว่าในสถานการณ์ใดที่อัลกอริทึมหนึ่งอาจมีประสิทธิภาพดีกว่าอีกอัลกอริทึมหนึ่ง (ยกตัวอย่างประกอบ)

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**9. Curse of Dimensionality (25 คะแนน)**

จงอธิบายแนวคิดของ "Curse of Dimensionality" ในบริบทของการเรียนรู้ของเครื่องว่าคืออะไร และส่งผลกระทบต่อประสิทธิภาพของโมเดลอย่างไร (เช่น K-Means หรือ k-NN) และอธิบายว่าเทคนิคการลดมิติข้อมูล (Dimensionality Reduction) เช่น PCA เข้ามาช่วยแก้ปัญหานี้ได้อย่างไร

**ชื่อ-นามสกุล:**........................................................................................ **รหัสนักศึกษา:**.............................................

**10. การตีความผลลัพธ์ (Result Interpretation) (25 คะแนน)**

สมมติว่าคุณได้สร้างโมเดล AI สำหรับตรวจคัดกรองโรคร้ายแรงที่พบได้ไม่บ่อย (เช่น พบผู้ป่วยเพียง 1% ของประชากร) หลังจากทดสอบโมเดลกับข้อมูลทดสอบ 1,000 คน พบว่าโมเดลมีความแม่นยำ (Accuracy) สูงถึง 99% จงวิเคราะห์ว่า:

10.1 ทำไมค่า Accuracy ที่สูงถึง 99% อาจไม่ใช่ตัวชี้วัดที่ดีที่สุดสำหรับปัญหานี้? (Hint: โมเดลอาจจะแค่ทายว่า "ไม่เป็นโรค" ทั้งหมด)

10.2 ระหว่างค่า Precision และ Recall ของคลาส "เป็นโรค" ตัวชี้วัดใดมีความสำคัญมากกว่ากันในสถานการณ์นี้ และเพราะเหตุใด?