1. จุดประสงค์หลักของการ Run คำสั่ง !chmod 600 ~/.kaggle/kaggle.json ใน Google Colab Notebook คืออะไร
    a. เพื่อแก้ไข Permission ของไฟล์
    b. เพื่อสร้าง Token ของ Kaggle API ใหม่
    c. เพื่อดาวน์โหลด Datasets จาก Kaggle
    d. เพื่อติดตั้ง Kaggle library

2. ควรเลือกใช้ระบบ CBM กับอุปกรณ์/เครื่องจักรประเภทใด?
    a. ใช้กับอุปกรณ์/เครื่องจักร ที่สามารถติดตั้งเซ็นเซอร์ได้ง่าย
    b. ใช้กับอุปรณ์/เครื่องจักรทุกประเภท
    c. ใช้กับอุปกรณ์/เครื่องจักรที่สำคัญที่สุด
    d. ใช้กับอุปกรณ์/เครื่องจักร ที่ใช้คนเข้าไปตรวจสอบได้ยาก

3. กระบวนการ Word-Embedding มีเทคนิดอะไรบ้าง?
    a. Skip-gram, CBOW, GloVe, FastText
    b. CNN, RNN, LSTM, Transformer
    c. TF-IDF, Bag-of-Words, N-gram, One-hot encoding
    d. PCA, t-SNE, UMAP, LDA

4. ถ้ามีการเรียกใช้ API สำหรับ สร้าง user ใหม่และผลของการทำงาน API สำเสร็จ ตัว API ควรจะ return http response code เป็นเลขอะไร
    a. 202
    b. 201
    c. 203
    d. 200

5. คำสั่งใดคือคำสั่งในการ Download Dockecker image จาก registry
    a. docker run
    b. docker build
    c. docker ps
    d. docker pull

6. ข้อใดต่อไปนี้สามารถใช้ตรวจสอบว่ามีคนกำลังเดินอยู่ในวิดีโอได้
    a. ถูกทุกข้อ
    b. นำภาพในเฟรมก่อนหน้ามาลบกับเฟรมปัจจุบัน หากมีความแตกต่างแสดงว่ามีการเคลื่อนไหว
    c. ใช้ smoothed nonlinear filter เพื่อตรวจจับการเคลื่อนไหว
    d. ใช้ median filter เพื่อลดสัญญาณรบกวนและทำให้รูปร่างของคนชัดเจนขึ้น

7. โครโมโซม 01111011 และ 01001111 ไม่น่าจะไขว้เปลี่ยนได้ลูกหน้าตาอย่างไร
    a. 11110111
    b. 01101111
    c. 01000111
    d. 01111111

8. ฟังก์ชันที่ใช่ในการอ่านรูปภาพ คือฟังก์ชันใด
    a. ฟังก์ชัน cv2.resize()
    b. ฟังก์ชัน cv2.imrite()
    c. ฟังก์ชัน cv2.imshow()
    d. ฟังก์ชัน cv2.imread()

9. การเชื่อมต่อแบบใด ไม่จัดอยู่ในรูปแบบ Wireless Communication
    a. Bluetooth
    b. PLC
    c. Wifi
    d. Cellular

10. การควบคุมอัตราการผสมพันธุ์ (Crossover Rate) มีความสำคัญอย่างไร
    a. เพื่อควบคุมจำนวนโครโมโซมใหม่ในแต่ละรุ่น
    b. ถูกทุกข้อ
    c. เพื่อรักษาสมดุลระหว่างการสำรวจและการปรับปรุง
    d. เพื่อป้องกันการทำลายโครโมโซมที่ดี

11. ข้อใดกล่าวถึงข้อแตกต่างของแบบจำลอง Tacotron2 (Auto-regressive) และ FastPitch (Non-autoregresive) ได้ถูกต้อง
    a. โมเดล Tacotron2 จะใช้เวลาในการสร้างเสียงสังเคราะห์น้อยกว่า FastPitch
    b. โมเดล FastPitch จะใช้เวลาในการสร้างเสียงสังเคราะห์มากกว่า Tacotron2
    c. โมเดล FastPitch จะมีการทำนายความยาวของหน่วยเสียงก่อน
    d. โมเดล Tacotron2 สามารถปรับแต่งระดับความทุ่มแหลมของเสียงได้

12. Lexitron มีความสำคัญกับการประมวลผลทางภาษาศาศาสตร์อย่างไร?
    a. Lexitron เป็นโปรแกรมที่ใช้ในการสร้างข้อความอัตโนมัติโดยใช้เทคนิคการเรียนรู้ของเครื่อง
    b. Lexitron เป็นพจนานุกรมอิเล็กทรอนิกส์ที่ใช้สำหรับการแปลภาษาจากภาษาไทยเป็นภาษาอังกฤษและภาษาอังกฤษเป็นภาษาไทย
    c. Lexitron เป็นชุดข้อมูลขนาดใหญ่ที่ประกอบด้วยตัวอย่างข้อความจากสื่อต่าง ๆที่ใช่ในการฝึกโมเดลภาษา
    d. Lexitron เป็นเครื่องมือที่ใช้ในการวิเคราะห์โครงสร้างประโยคและการวิเคราะห์ความหมายของประโยคในภาษาต่าง ๆ

13. Cryptocurrency & Software Wallet aguah แต่ท่าไมต้องใช้ Hardware Wallet เพิ่มเติม ?
    a. Hardware Wallet ใช้งานง่ายกว่า Software Wallet
    b. Hardware Wallet พกพาสะดวกกว่า Software Wallet
    c. Hardware Wallet รองรับ Cryptocurrency มากกว่า Software Wallet
    d. Hardware Wallet มีความปลลดภัยสูงกว่า Software Wallet

14. เทคนิคในข้อใดไม่ได้ถูกประยุกต์ใช้กับการหา Word Similarity?
    a. Pointwise Mutual Information (PMI)
    b. Cosine Similarity
    c. Euclidean Distance
    d. Random Forest

15. ในขณะที่หุ่นยนต์ส่งสินค้าเดินไปตามทางเท้าที่มีคนพลุกพล่าน มีสิ่งก็ดขวางเข้ามาขวางด้านหน้าหุ่นยนต์อย่างกระทันหันสถานการณ์ใดต่อไปนี้ ที่แสดงให้เห็นถึง การทำงานตามกระบวนการ Sense-Think-Act ของหุ่นยนต์ได้ดีที่สุด
    a. หุ่นยนต์หยุดกะทันหัน ทำให้คนที่เดินอยู่ด้านหลังชนแล้วล้ม
    b. หุ่นยนต์คำนวณหาเส้นทางใหม่ รอบ ๆ สิ่งกีดขวาง โดยใช้เวลานานเป็นสองเท่าในการไปถึงจุดหมายปลายทาง
    c. หุ่นยนต์ตรวจจับสิ่งกีดขวาง แต่ยังคงไม่หยุดเคลื่อนที่ไปข้างหน้า
    d. หุ่นยนต์ส่งเสียงเตือนเสียงดัง สร้างความตกใจให้กับคนที่เดินอยู่บริเวณนั้น

16. ในกรณีของ DeepFake นักศึกษาคิดว่า มีความเสี่ยงที่จะละเมิดหลักการของปัญญาประดิษฐ์ในข้อใดมากที่สุด
    a. ความเท่าเทียม หลากหลาย ครอบคลุม และเป็นธรรม
    b. ความมั่นคงปลอดภัยและความเป็นส่วนตัว
    c. ความสามารถในการแข่งขันและพัฒนาอย่างยั่งยืน
    d. ความน่าเชื่อถือ

17. ข้อใดถูกต้องเกี่ยวกับ Max Pooling?
    a. Max Pooling เป็นกระบวนการที่ใช้ในการทำให้ข้อมูลนำเข้าเป็นเชิงเส้น
    b. Max Pooling เป็นกระบวนการที่ใช้ในการเพิ่มขนาดของข้อมูลนำเข้าโดยการเพิ่มข้อมูลซ้ำในแต่ละกลุ่มของข้อมูล
    c. Max Pooling เป็นกระบวนการที่ใช้ในการสกัดข้อมูลที่มีค่าเฉลี่ยสูงสุดในแต่ละกลุ่มของข้อมูล
    d. Max Pooling เป็นกระบวนการที่ใช้ในการลดขนาดของข้อมูลนำเข้าโดยการเลือกค่ามากสุดในแต่ละกลุ่มของข้อมูล

18. ระบบพิกัดในข้อใดที่ปกติจะอยู่บนระนาบเดียวกัน
    a. Image pixel coordinate system และ Camera coordinate system
    b. Image pixel coordinate system และ  Imaging plane coordinate system
    c. Imaging plane coordinate system และ World coordinate system
    d. Camera coordinate system และ Imaging plane coordinate system

19. สถานการณ์ใดต่อไปนี้ แสดงให้เห็นถึงการใช้ Automatic completion ของ Google Colab ได้มีประสิทธิภาพที่สุด
    a. พิมพ์ชื่อ Function และ Parameter แบบเต็มด้วยตัวเองทั้งหมด
    b. ใช้ Autocomplete สำหรับ Function ทั่วไป จากนั้นปรับแต่งตามต้องการ
    c. กด โลb ทุกครั้ง หลังจากพิมพ์ตัวอักษร ไปได้ 1 ตัวอักษร
    d. ปิดการใช้งาน Autocomplete ทั้งหมดเพื่อหลีกเลี่ยงการรมกวน

20. ข้อใดคือความสัมพันธ์ระหว่าง exploration และ exploitation
    a. exploration และ exploitation มากเกินไปจะทำให้อัลกอริธึมทำงางานช้าลง
    b. exploration และ exploitation มากเกินไปจะทำให้อัลกอริธึมไม่สามารถหาผลลัพธ์ได้เลย
    c. exploration มากเกินไปอาจทำให้อัลกอริธึมติดอยู่ในผลลัพธ์ท้องถิ่น ในขณะที่ exploitation มากเกินไปอาจทำให้พลาดโซลูชั่นที่ดีที่สุด
    d. exploration มากเกินไปอาจทำให้อัลกอริธึมพลาดโซลูชันที่ดีที่สุด ในขณะที่ exploitation มากเกินไปอาจทำให้อัลกอริธึมติดอยู่ในผลลัพธ์ท้องถิ่น

21. ข้อใด กล่าวถึง โครงสร้างโปรแกรมภาษา C ที่ใช้กับ Arduino ได้ถูกต้อง
    a. ฟังก์ชัน initialize() ฟังก์ชัน run()
    b. ฟังก์ชัน setup() ฟังก์ชัน loop()
    c. ฟังก์ชัน start) ฟังก์ชัน repeat()
    d. ฟังก์ชัน begin) ฟังก์ชัน cycle()

22. ในการสังเคราะห์เสียงพูดจะต้องอาศัย Vocoder เพื่อแปลงจากผลลัพธ์ที่ได้จากโมเดล Tacotron หรือ FastPitch เป็นเสียงพูด ข้อใดเป็นข้อมูลนำเข้าโมเดล Vocoder
    a. F0
    b. BAP
    c. Mel-spectrogram แบบไม่มี phase informaton
    d. MFCC

23. Amazon เก็บข้อมูลผู้บริโภคจำนวนวนมหาศาลเพื่อวิเคราะห์พฤติกรรมการซื้อและค้นหาสินค้า โมเดลข้อมูลใดต่อไปนี้ ไม่ใช่โมเดลหลักที่ Amazon ใช้เพื่อวิเคราะห์พฤติกรรมของผู้บริโภค ?
    a. โมเดลข้อมูลลูกค้า (Customer Data Model) โมเดลนี้เก็บข้อมูลส่วนตัว ข้อมูล การติดต่อ ประวัติการซื้อ และข้อมูลอื่นๆ เกี่ยวกับลูกค้าแต่ละราย
    b. โมเดลข้อมูลสินค้า (Product Data Model) โมเดลนี้เก็บข้อมูลเกี่ยวกับสินค้าที่ Amazon ขาย เช่น รายละเอียดสินค้า ราคา รูปภาพ และรีวิว
    c. โมเดลข้อมูลการจัดส่ง (Delivery Data Model) โมเดลนี้เก็บข้อมูลเกี่ยวกับที่อยู่จัดส่ง วันที่จัดส่ง และข้อมูลติดตามสินค้า
    d. โมเดลข้อมูลการค้นหา (Search Data Model) โมเดลนี้เก็บข้อมูลเกี่ยวกับคำค้นหาของลูกค้า สินค้าที่พวกเขาเลือกใส่ตะกร้า และสินค้าที่พวกเขาซื้อ

24. เหตุใดเมื่อน่า PCA ไปประยุคต์ใช้ในการรู้จำใบหน้าภายใต้สภาวะแสงที่หลากหลาย เมื่อการเลือกค่า eigenvectors ที่สอดคล้องกับค่า eigenvalue ต่ำไปใช้แล้วกลับได้ประสิทธิภาพสูง
    a. เนื่องจากองค์ประกอบที่ได้จากค่า eigenvalue สูง มีค่า Mean ของแสงมาก
    b. เนื่องจากองค์ประกอบที่ได้จากค่า eigenvalue สูง มีค่า Variance ของแสงมาก
    c. เนื่องจากองค์ประกอบที่ได้จากค่า eigenvalue สูง ทำให้ลด Dimension ได้มาก
    d. เนื่องจากองค์ประกอบที่ได้จากค่า eigenvalue สูง มีค่า Noise มาก

25. ข้อใดไม่ถูกต้องในการใช้งาน L1 และ L2 loss function
    a. ค่าความคลาดเคลื่อนสามารถติดลบได้
    b. ใช่ในการหาค่าพารามิเตอร์ที่เหมาะสม
    c. ใช่ในการบอกทิศทางในการปรับค่าของพารามิเตอร์
    d. ค่าความคาดเคลื่อนต้องเป็นค่าติดลบ

26. การหาพื้นที่ของวัตถุเป้าหมายในภาพ จัดเป็นงานประเภทใดในด้านComputer Vision
    a. Regression (การประมาณค่า)
    b. Classification (การจำแนกประเภท)
    c. Segmentation (การแบ่งส่วนภาพ)
    d. Object Detection (การตรวจจับวัตถุ)

27. จุดกำเนิด (0,0) ของระบบพิกัด Image pixel coordinate systemอยู่ในตำแหน่งใด
    a. ขวาบน
    b. ขวาล่าง
    c. ซ้ายบน
    d. ซ้ายล่าง

28. วิธีใดช่วยจัดการคำที่ไม่เคยปรากฏ (unknown word) ในการแปลได้ที่ดีที่สุด
    a. ลบคำนั้นออกจากประโยคภาษาต้นทาง
    b. แบ่งข้อความเป็นระดับตัวอักษรแล้วสร้างโมเดลแปลระดับตัวอักษร
    c. แทนคำคำที่ไม่ปรากฏในคลังข้อมูลด้วยสัญสัญลักษณ์ บnk
    d. แบ่งข้อความเป็นหน่วยย่อยที่เล็กกว่าระดับคำ (Subword unit) แต่ใหญ่กว่าระดับตัวอักษร

29. ในการทดสอบระบบตรวจจับความผิดปกติของการจราจรทางอากาศ ผลลัพธ์จาก Confusion Matrix แสดงให้เห็นว่ามี False Alarm (False Positive) สูงมาก uain Miss (False Negative) Rin จากสถานการณ์นี้ ข้อใดเป็นการแปลความหมายและข้อเสนอแนะที่เหมาะสมที่สุด
    a. ระบบทำงานได้ดีแล้ว เพราะสามารถตรวจจับความผิดปกติได้เกือบทั้งหมด ไม่จำเป็นต้องปรับปรุง
    b. ควรเปลี่ยนไปใช้ deep learning-based approach เพราะวิธีปัจจุบันไม่สามารถจัดการกับความซับซ้อนของข้อมูลได้
    c. ควรเพิ่มจำนวนข้อมูล normal ในชุดชุดข้อมูลฝึกฝน เพื่อให้ระบบเรียนรู้รูปแบบปกติได้ดีขึ้น
    d. ระบบมีความไว (Sensitivity) สูงเกินไป ควรปรับลดความไวลงเพื่อลด FalseAlarm แม้อาจทำให้ Miss เพิ่มขึ้นบ้าง

30. Softmax คืออะไร และถูกนำไปไปใช้งานอย่างไรในวันวิทยาการคอมพิวเตอร์?
    a. Softmax เป็นอัลกอริทึมสำหรับการลดขนาดของข้อมูล โดยใช่ในการประมวลผลภาพและเสียง ใช้ในงานลดข้อมูลรบกวนและ feature extraction
    b. Softmax เป็นฟังก์ชั่นที่ใช้ในการวัดความคล้ายคลึงกันระหว่างเวกเตอร์สองตัวใช่ในงานค้นหาข้อมูลและการแนะนำ
    c. Softmax เป็นฟังก์ชันที่ใช้แปลงค่าจำนวนจริงชุดหนึ่งให้เป็นค่าความน่าจะเป็นโดยค่าทั้งหมดจะรวมกันได้ 1 และค่าที่มากที่สุดจะสอดคล้องกับคลาสที่มีความน่าจะเป็นสูงสุด ใช่ในงานจำแนกประเภท เช่น การจำแนกภาพหรือข้อความ
    d. Softmax เป็นวิธีการหนึ่งในการปรับปรุงประสิทธิภาพของ neural network โดยใช่ในการปรับค่าน้ำหนักของแต่ละ node ใน network

31. ข้อใดอธิบายการใช้ประโยชน์ของ Eigenvalues และ Eigenvectors ในการวิเคราะห์ข้อมูลได้ถูกต้อง?
    a. ใช่ในการเปลี่ยนข้อมูลเชิงเส้นให้เป็นข้อมูลไม่เชิงเส้น
    b. ใช้ในการหาค่าที่ต่ำสุดของเมทริกซ์เพื่อลดความซับซ้อนของข้อมูล
    c. ใช้ในการทำให้ข้อมูลที่มีมิติสูงขึ้นเพื่อการวิเคราะห์ที่ซับซ้อนขึ้น
    d. ใช้ในการหาองค์ประกอบที่สำคัญในข้อมูลโดยกระบวนการ PCA

32. Observation probability คือ ค่า probablity ของตัวแปร a11, a12, a22, a23, a33
    a. True
    b. False
33. ข้อใดต่อไปนี้กล่าวไม่ถูกต้องเกี่ยวกับ Feature descriptor
    a. Local feature descriptor จะน่าจุด 3 จุดที่อยู่ดำแหน่งรอบๆ ของจุด Keypoint มาใช้พิจารณาลักษณะของจุด Keypoint
    b. Feature descriptor ที่ดีต้องยังคงบอกคุณสมบัติลักษณะของจุด Keypoint ได้ แม้ค่า Resolution จะเปลี่ยนไป
    c. Feature descriptor ที่ดีต้องยังคงบอกคุณสมบัติลักษณะของจุด Keypoint ได้ แม้จะมีสัญญาณรบกวนของภาพ
    d. Global feature descriptor จะน่าจุดใน Point cloud ทุกจุดมาใช้พิจารณาลักษณะของจุด Keypoint

34. POS tagging คืองานในข้อใด
    a. การระบุขอบเขตของนิพจน์ระบุนาม
    b. การกำกับชนิดของคำ
    c. การวิเคราะห์โครงสร้างประโยค
    d. การตีความความหมายแฝงของคำ

35. ข้อใดกล่าวผิดเกี่ยวกับการใช้งาน HTTP Method ร่วมกับ RESTful API
    a. การเปลี่ยนแปลงข้อมูลใช้งาน HTTP Method PUT หรือ PATCH
    b. การเพิ่มข้อมูลควรใช้งาน HTTP Method POST
    c. การอ่านข้อมูลควรใช้งาน HTTP Method GET
    d. การลบข้อมูลควรใช้งาน HTTP Method DELETE

36. วิธีการใดที่ Flatten histogram ได้มากที่สุด
    a. Local intensity distribution equalization
    b. Histogram equalization
    c. Contrast Limited Adaptive Histogram equalization
    d. Pixel sorting

37. Algorithm ใดต่อไปนี้สามารถน่ามาใช้ในการแก่ไขปัญหาเรื่อง
    a. missing value ในข้อมูลแบบต่อเนื่อง หรือในข้อมูลแบบกลุ่มได้
    b. Linear Regression
    c. K-NN
    d. Logistic Regression

38. ในกระบวนการแปลงภาพเป็นภาพดิจิทัล Sampling หมายถึงข้อใด
    a. การเพิ่มความคมชัดของภาพโดยการใช้ Fiter ที่เหมาะสม
    b. กระบวนการกำหนดตำแหน่งข้อมูลบนภาพที่ต่อเนื่องเป็นข้อมูลแบบไม่ต่อเนื่อง เพื่อบันทึกค่าแสง
    c. การแปลงค่าสีของภาพที่มีค่าต่อเนื่องให้เป็นระดับสีเพื่อบันทึกในดอมพิวเตอร์
    d. การลดสัญญาณรบกวนในภาพโดยการกำจัดจุดดำและจุดขาวที่ไม่ต้องการ

39. จงจับคู่คำที่มีความเกี่ยวข้องกันต่อไปนี้ให้ถูกต้อง

- A.
    - MySQL
    - MSSQL
- B.
    - Cassandra
    - HBase
- C.
    - Memcached
    - Coherence

- D.
    - Neo4J
    - OrientDB
- E.
    - MongoDB
    - CouchDB

ประเภทฐานข้อมูลที่ต้องจับคู่:

1. Graph DB
2. Document DB
3. Key-Value Cache
4. RDBMS
5. Tabular/Columnar DB

ตัวเลือกคำตอบ:

a. C, B, D, A, E
b. D, C, E, A, B
c. B, C, D, A, E
d. D, E, C, A, B

40. ภาษาใดต่อไปนี้ใช้ระบบการเขียนแบบ Alphabets
    a. ภาษาพม่า, ภาษาไทย, ภาษาทมิฬ
    b. ภาษาอารบิก, ภาษาฮิบรู, ภาษาแอราเมอิก
    c. ภาษาโรมัน, ภาษาอาร์มีเนีย, ภาษากรีก
    d. ภาษาเชโรกี, อักษรครี, ฮิรางานะ(ญี่ปุ่น)

41. ข้อใดต่อไปนี้กล่าวไม่ถูกต้อง
    a. Feature matching สามารถเปรียบเทียบกันโดยใช้ Histogram ของ Keypointที่ต้องการ
    b. Keypoint ใน Point cloud หนึ่งสามารถตรวจสอบว่าเป็น Keypoint ในอีก Point cloud หนึ่งได้โดยใช้ Transformation matrix ในการคำนวณ
    c. Keypoint ทุกคู่ที่ได้จากการดำเนินการ Matching จะถูกนำมาใช่ในการค่านวณหา Transformation matrix
    d. Point matching สามารถเปรียบเทียบกันโดยใช้พิกัด 3 มิติของ Keypoint ที่ต้องการ

42. วิธีการใดสามารถจัดการ Noise image ได้
    a. Spatial Filtering
    b. Segmentation
    c. Image sampling
    d. Image quantization

43. กระบวนการ Stemming และ Lemmatization เหมือนหรือแตกต่างกันอย่างไร
    a. Stemming และ Lemmatization เป็นกระบวนการเดียวกัน ใช้เพื่อลดรูปคำให้เป็นรากศัพท์
    b. Lemmatization เป็นกระบวนการที่รวดเร็วกว่า Stemming แต่ให้ผลลัพธ์ที่แม่นยำน้อยกว่า
    c. Stemming ใช้พจนานุกรมในการลดรูปคำ ส่วน Lemmatization ใช้กฎทางภาษาศาสตร์
    d. Stemming ตัดส่วนท้ายของคำออก ในขณะที่ Lemmatization ใช้ความรู้ทางภาษาศาสตร์เพื่อลดรูปคำให้ถูกต้องตามหลักไวยากรณ์

44. Category Comparison เหมาะสำหรับการแสดงข้อมูลแบบใด
    a. ข้อมูลที่ต้องการแสดงความสัมพันธ์
    b. ข้อมูลที่มีการเชื่อมโยงกัน
    c. ข้อมูลที่ต้องการแสดงความถี่
    d. ข้อมูลที่มีการเปรียบเทียบ

45. ในการใช้งาน Google Colab สำหรับโปรเจค Machine Learning ที่ต้องการประสิทธิภาพการประมวลผลสูง หากต้องการตรวจสอบว่า GPU ที่ได้รับการจัดสรรมีสเปคอย่างไร คำสั่งใดต่อไปนี้เหมาะสมที่สุดสำหรับการตรวจสอบนี้
    a. from google.colab import files
    b. print(tf.test.gpu_device_name())
    c. !nvidia-smi
    d. !pip freeze

46. ภาษาไทยใช้ระบบการเขียนแบบใด?
    a. Abjads
    b. Alternative scripts
    c. Abugidas
    d. Alphabets

47. โค้ดข้อใดเป็นคำสั่งที่ใช้แบ่งข้อมูลลออกเป็น 2 ส่วน โดยให้ส่วนข้อมูลเทรนมีขนาดเป็น 80% ของข้อมูลทั้งหมด
    a. from sklearn.model selection import train_test_split X_train, y_train, X_test, y_test = train_test_split(X, y, test_size=0.80)
    b. from sklearn.model_selection import train_test_split X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.20)
    c. from sklearn.model_selection import train_test_split X_train, y_train, X_test, y_test = train_test_split(X, y, train_size=0.80)
    d. from sklearn.model selection import train test_split X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)

48. หากข้อมูลน่าเข้าเป็นรูปภาพสี RGB ขนาด 225 x 225 ถูกประมวลผลด้วย Convolution layer ที่มีขนาด 3 3 3 (width x height) จำนวน 64 filters โดยใช้ Stride=1 ไม่ใช้ Padding ขนาดของผลลัพธ์ของแต่ละ filter ที่ได้จะมีจำนวนเท่ากับข้อใด?
    a. 223 x 223 x 128
    b. 223 x 223 x 1
    c. 223 x 223 x 64
    d. 223 x 223 x 3

49. การทำ ฉันทมติ (Consensus) ด้วยวิธีการ PoS ทำไมถึงประหยัด พลังงานมากกว่าวิธีการ PoW ?
    a. PoS ใช้การคำนวณทางคณิตศาสตร์ที่เรียบง่ายกว่า PoW
    b. POS ไม่จำเป็นต้องใช้ฮาร์ดแวร์พิเศษสำหรับการขุด
    c. PoS ตรวจสอบธุรกรรมโดยใช้เครือข่ายผู้ตรวจสอบแบบกระจายศูนย์
    d. PoS จูงใจให้ผู้ใช้ถือเหรียญไว้แทนที่จะขุด

50. ข้อความต่อไปนี้ เกี่ยวข้องกับ Data Types และ Database จงเลือกคำให้เหมาะสมกับข้อความเหล่านี้ " โครงสร้างเรียบง่าย มนุษย์สามารถอ่านและเขียนได้โดยไม่ต้องใช้เครื่องมือพิเศษ ใช้พื้นที่เก็บข้อมูลน้อยเหมาะสำหรับการส่งผ่านข้อมูลบนอินเทอร์เน็ต รองรับภาษาโปรแกรมหลากหลายภาษา เหมาะสำหรับการพัฒนาเว็บแอปพลิเคชัน เป็นภาษาที่ถูกออกแบบมาเพื่อจัดการกับฐานข้อมูลโดยเฉพาะ ใช้งานได้กับระบบฐานข้อมูลแบบเชิงสัมพันธ์ เรียนรู้และใช้งานได้ง่าย โดยใช้คำศัพท์ภาษาอังกฤษทั่วไปถูกออกแบบบมาเพื่อรองรับข้อมูลที่มีรูปแบบหลากหลาย เหมาะส่าหรับข้อมูลที่ไม่มีโครงสร้างตายตัว หรือโครงสร้างที่เปลี่ยนแปลงอยู่เสมอ และสามารถปรับขนาดได้ง่ายเหมาะสำหรับการใช้งานกับแอปพลิเคชันสมัยใหม่ที่มีการใช้งานข้อมูลจำนวนมาก"
    a. JSON,JSON,SQL
    b. JSON,SQL,NoSQL
    c. SQL,JSON,SQL
    d. SQL,JSON,NoSQL

51. เหตุใดการอัพเดทความรู้ใน Bayesian inference จึงมีประโยชน์?
    a. เพราะช่วยในการสร้างข้อมูลใหม่จากความรู้เดิม
    b. เพราะช่วยในการเพิ่มความน่าจะเป็นของเหตุการณ์ที่ไม่เกิดขึ้น
    c. เพราะช่วยในการลดความซับซ้อนของปัญหา
    d. เพราะช่วยในการปรับปรุงความน่าจะเป็นของเหตุการณ์ตามข้อมูลใหม่

52. สมมติว่าเราต้องการ maximize function เราต้องทำเช่นไร
    a. ใช้ Viterbi algorithm
    b. ทำ eigen-decomposition
    c. adapt ตัวแปรสวนทิศ gradient
    d. adapt ตัวแปรตามทิศ gradient

53. การทำ Hash แบบ SHA256 ดีกว่าการทำ Hash แบบ Encryption/Decryption arinols ?
    a. Hash แบบ SHA256 ตรวจจับการเปลี่ยนแปลงข้อมูลได้ดีกว่า Encryption/Decryption
    b. Hash แบบ SHA256 เร็วกว่า Encryption/Decryption
    c. Hash แบบ SHA256 ใช้พลังงานในการประมวลผลน้อยกว่า Encryption/Decryption เสมอ
    d. Hash แบบ SHA256 ปลอดภัยกว่า Encryption/Decryption เสมอ

54. สมมติว่า H0 ใน ANOVA ถูกปฏิเสธ หมายความว่าอย่างไร?
    a. ค่าเฉลี่ยของทุกกลุ่มเท่ากัน
    b. มีความแปรปรวนของกลุ่มหนึ่งที่แตกต่างจากความแปรปรวนของกลุ่มอื่นๆ
    c. มีค่าเฉลี่ยของกลุ่มหนึ่งที่แตกต่างจากค่าเฉลี่ยของกลุ่มอื่นๆ
    d. ความแปรปรวนของทุกกลุ่มเท่ากัน

55. หากต้องการดู log การทำงานภายใต้ container id a036be714862 ที่ถูกสร้างให้ท่างานด้วยคำสั่ง docker run -P nginx:alpine จะสามารถดูได้โดยใช้คำสั่งใด
    a. ไม่ต้องทำอะไรเพราะ log จะแสดงที่หน้าจอโดยอัดโนมัติ
    b. docker log nigix:alpine
    c. docker log a036be7f4862
    d. docker logs a036be7f4862

56. การพัฒนาระบบตรวจจับความผิดปกติสำหรับการเคลื่อนที่ของเรือ พบว่าระบบมีแนวโน้มที่จะระบุเรือที่เคลื่อนที่ช้ามากหรือเร็วมากเกินไปเป็น outier บ่อยครั้ง แม้ว่าผู้เชี่ยวชาญจะระบุว่าเป็นการเคลื่อนที่ปกติ จากสถานการณ์นี้ ข้อใดเป็นแนวทางที่เหมาะสมที่สุดในการปรับปรุงประสิทธิภาพของระบบ
    a. เพิ่มมิติของข้อมูลโดยรวมปัจจัยอื่น เช่น ขนาดของเรือและสภาพอากาศ ในการวิเคราะห์
    b. เพิ่มจำนวนข้อมูลการเคลื่อนที่ของเรือที่มีความเร็วป่านกลางในชุดข้อมูลฝึกฝน (training data)
    c. ปรับเปลี่ยนค่า threshold ของ Standard Deviation ในการระบุ outlier จาก30 เป็น 20
    d. ใช้เทคนิค deep learning แทนวิธีการแบบดั้งเดิมในการตรวจจับ outlier

57. ข้อใดไม่ใช่องค์ประกอบของห่วงโซ่คุณค่าของโมเดลธุรกิจข้อมูล
    a. Big data Application Provider/Big Data Framework Provider
    b. Data Consumer/Data Provider
    c. System Orchestrator
    d. Engineering/Governance

58. ข้อใด คือ องค์ประกอบหลักของ IoT
    a. Internet, Microcontroller, 5G
    b. Hardware, Software, Connectivity
    c. Memory, Input, Output
    d. Lora, Sigfox, NB-IoT

59. Library ใดที่สามารถช่วยในการสร้าง Decision Tree ได้
    a. numpy
    b. pandas
    c. opencv
    d. scikit-learn

60. ในการพัฒนาโมเดล AI เพื่อวิเคราะห์ความคิดเห็นของลูกค้ารีวิวสินค้าออนไลน์ ประเภทของการเรียนรู้ของเครื่อง (Machine Learning) ที่เหมาะสมที่สุดคืออะไร
    a. Supervised Learning
    b. Semi-supervised Learning
    c. Unsupervised Learning
    d. Reinforcement Learning

61. การมีระบบ IoT ไม่ได้ช่วยผลักดันในเรื่องใด
    a. ค่าใช้จ่าย
    b. ประสิทธิภาพ
    c. การใช้ไฟฟ้า
    d. นวัตกรรม

62. คำสั่ง sudo ใช้เพื่ออะไร?
    a. เพื่อแจ้งให้ Superuser ทำงานให้เรา
    b. เพื่อทำงานโดยใช้สิทธิ์ของ Superuser
    c. เพื่อดูคิวงานที่ Superuser ต้องทำ
    d. เพื่อทำงานใน Secure user mode

63. ข้อใดจัดเป็นความแตกต่างระหว่าง Adaptive histogram equalization hu Local intensity distribution equalion
    a. Global vs local
    b. Gray vs RGB
    c. Single vs multivariable
    d. Non-parametric vs parametric

64. Decision Tree ใช้ค่าใดในการกำหนด Feature เพื่อ Split
    a. Mean Squared error
    b. ถูกทุกข้อ
    c. Entropy
    d. Mean Absolute error

65. ให้ผลลัพธ์การแปลคือ Target sentence คือ "ฉัน กิน ข้าว อยู่ ใต้ต้นไม่" ประโยคที่เป็น reference คือ ฉัน กิน ขนม อยู่ ใต้ ต้นไม้ในการคำนวณ BLEU Score จำเป็นต้องคำนวณ Cumulative n-gramจากตัวอย่าง ดังกล่าวจำนวน n และค่า Cumulative 1-gram ถึง n-gram คือเท่าไร
    a. n = 6, Cumulative 1-gram ถึง n-gram 5/6,3/5, 1/4,0, 0, 0, 0
    b. n = 5, Cumulative 1-gram ถึง n-gram 5/6, 3/5, 1/4, 0, 0, 0
    c. n = 6, Cumulative 1-gram ถึง n-gram 5/6, 3/4, 1/4, 0, 0, 0
    d. n = 4, Cumulative 1-gram ถึง n-gram 5/6,2/5, 1/4,0, 0, 0, 0

66. นักศึกษาทุกคนที่เรียน superAI แล้วจะเก่ง แปลได้ว่า
    a. Vx( study(x, superAI) → smart(x) )
    b. ∃x( study(x, superAI) → smart(x) )
    c. ∃x( study(x, superAI) v smart(x) )
    d. Vx( study(x, superAI) A smart(x) )

67. จากโค้ดต่อไปนี้ คำสั่ง layers.Flatten ทำหน้าที่อะไร

```
from tensorflow.keras.applications.vgg16 import VGG16
vgg = VGG16(include_top=False, weights='imagenet')
# fit input
x_in = layers.Input(shape=(32, 32, 1))
x = layers.Conv2D(3, 1)(x_in)
x = vgg (x)
# fit output
x = layers.Flatten()(x)
x = layers.Dense(10, activation-'softmax') (x)
model = keras.Model (x_in, x)
model.summary ()
```

    a. แปลงข้อมูลจาก multi-dimensions ให้เป็น vector
    b. เพิ่มจำนวนช่องสีให้เข้ากับ weights ที่ใช้งาน
    c. แปลงข้อมูล multi-layers ให้เป็น a layer
    d. แปลงข้อมูลให้มีจำนวน output เท่ากับ 10

68. ข้อใด ไม่จัดเป็นอุปกรณ์สำหรับใช้ใช่ในการพัฒนา IoT
    a. Arduino Board
    b. ESP8266
    c. Mainboard
    d. Raspberry Pi

69. ข้อใดกล่าวถึง Covariance matrix ได้ถูกต้องมากที่สุด
    a. ค่า Covariance เท่ากับค่า Variance และเท่ากับค่า Standard Deviation
    b. ขนาดของ Covariance matrix เป็นเมทริกซ์จัดรัสเสมอ
    c. Covariance(x,x) ไม่เท่ากับ Variance(x)
    d. Covariance(x,y) ไม่เท่ากับ Covariance(y,x)

70. การแทนลำดับข้อมูลนำเข้า (input sequence) ในข้อใด หากนำไปสร้างระบบสังเคราะห์เสียง แล้วจะมีความถูกต้องและเป็นธรรมชาติมากกว่าข้ออื่น
    a. ตัวอักษร (Character)
    b. จำนวนของพยางค์ในแต่ละคำ
    c. สัญลักษณ์แทนการออกเสียง (Phoneme)
    d. ชนิดของคำ (Part-of-speech)

71. ข้อใดเป็นเหตุผลที่ Perceptron ไม่สามารถแก้ปัญหา XOR ได้?
    a. เนื่องจากใช้ฟังก์ชันกระตุ้นแบบ ReLU
    b. เนื่องจากการปรับค่า Hyperparameters ที่ไม่เหมาะสม
    c. Perceptron ไม่สามารถเรียนรู้รูปแบบของข้อมูลแบบเชิงเส้นได้
    d. XOR เป็นปัญหาไม่เชิงเส้น แต่ Perceptron แก่ได้เฉพาะปัญหาเชิงเส้น

72. Lie Factor จากหัวข้อ Graphical Integrity ควรอยู่ในช่วงเท่าใด
    a. 0.80 - 1.20
    b. 0.90 - 1.10
    c. 0.85 - 1.15
    d. 0.95 - 1.05

73. DOF ของระบบแขนกลหุ่นยนต์เรียกว่าชื่อเต็มว่าอย่างไร
    a. Direct of Freedom
    b. Degree of Freedom
    c. Degree of Field
    d. Direct of Factor

74. วิธีการ Vanilla ICP นั้นประกอบด้วยขั้นตอนตามลำดับอย่างไรบ้าง
    a. Alignment was Data Association
    b. Distance measuring และ Alignment
    c. Data association แล: Alignment
    d. Matrix decomposition และ Data association

75. เมื่อไม่ได้กำหนดฟังก์ชั่นกระตุ้น (Activation) ให้กับชั้นต่าง ๆ หมายความได้ว่าเป็นการใช้ฟังก์ชันกระตุ้นแบบเชิงเส้น (Linear)

1. TRUE
2. FALSE

76. ข้อใดอธิบายการทำงานของ Bootstrap sampling ได้ถูกต้อง

a. sample without replacement
b. sample with replacement multiple times
c. without sampling
d. sample with replacement

77. ข้อใดเรียงลำดับการประมวลผลของการรู้จำเสียงพูดได้ถูกต้อง

a. Signal level - Acoustic level - Language Level
b. Signal level - Language Level - Acoustic level
c. Language Level - Acoustic level - Signal level
d. Acoustic level + Language Level - Signal level

78. โมเดล U-Net ต่างจาก SegNet อย่างไร

a. U-net มีการลดจำนวน layer ใน Decoder เพื่อให้ท่างางานเร็วขึ้น
b. U-net มีการลดจำนวน layer ใน Encoder เพื่อให้ทำงานเร็วขึ้น
c. U-net มีการเดิมข้อมูลจาก Decoder ใส่ Encoder เพื่อเพิ่มข้อมูลที่คาดว่าจะหายไป
d. U-net มีการเติมข้อมูลจาก Encoder ใน Decoder เพื่อเพิ่มข้อมูลที่คาดว่าจะหายไป

79. Self-Attention ช่วยในการจัดการกับลำดับที่ยาวได้อย่างไร

a. โดยการลดขนาดของลำดับข้อมูลให้สั้นลง
b. โดยการคำนวณความสำพันธ์ระหว่างข้อมูลทุกตำแหน่งในล่าดับพร้อมกัน
c. โดยการใช้ Recurrent Connections เพื่อเพิ่มหน่วยความจำ
d. โดยการใช้ Convolutional Layers เพื่อจับลักษณะของล่าดับ

80. ข้อใดไม่ใช่ประโยชน์ของการใช้ LDA topic modeling

a. การค้นพบ themes ที่ซ่อนอยู่ในคลังข้อมูลขนาดใหญ่
b. การจัดหมวดหมู่เอกสารอัตโนมัติ
c. การปรับปรุงระบบค้นคืนสารสนเทศ
d. การสรุปเนื้อหาของเอกสาร
e. การแก้ไขข้อผิดพลาดทางไวยากรณ์ในเอกสาร

81. ข้อในต่อไปนี้กล่าวไม่ถูกต้องเกี่ยวกับ Independent Component Analysis (ICA)
    a. ใช้สำหรับการลดขนาดมิติของข้อมูล
    b. ICA สามารถนำมาช่วยในการหาโครงสร้างของข้อมูล
    c. ใน ICA แต่ละ component มีความสำคัญไม่เท่ากัน
    d. ใช้สำหรับการแยกคุณลักษณะเฉพาะ

82. ข้อใดเป็นเป้าหมายหลักของการควบคุมแรงกดดันการคัดเลือก(Selection Pressure)
    a. เพื่อหาคำตอบที่เหมาะสมที่สุดอย่างมีประสิทธิภาพ
    b. เพื่อเพิ่มความหลากหลายในประชากร
    c. ถูกทุกข้อ
    d. เพื่อเร่งกระบวนการวิวัฒนาการ

83. ข้อใดไม่ใช่ประเภทของ Influencer
    a. Executive Influencer
    b. Niche Influencer
    c. Virtual Influencer
    d. Exclusive Influencer

84. ข้อใดกล่าวไม่ถูกต้องเกี่ยวกับ Histogram
    a. การทำ Histogram equalization จะทำให้ Histogram มีช่วงการกระจายกว้างขึ้น
    b. Histogram คือกราฟที่แสดงจำนวนวัตถุของภาพ
    c. การทำ Histogram equalization จะทำให้ภาพมีความต่างสีกันมากขึ้น
    d. ภาพที่มีสีแบบ Low contrast นั้น Histogram จะมีช่วงการกระจายในระยะแคบ

85. ข้อมูลชนิดใดที่สามารถนำไปสร้างเป็นภาพได้
    a. Float64
    b. ถูกทุกข้อ
    c. Float32
    d. UINT8

86. ข้อใดไม่ถูกต้อง เกี่ยวกับการแข่งขันแบบ Team ใน Kaggle?
    a. ชื่อทีม สามารถเปลี่ยนได้ตามความต้องการ
    b. แรกเริ่ม ทุกคนจะอยู่ในทีม ที่เป็นชื่อของตัวเอง
    c. ในการแข่งแบบทีม เราสามารถส่งคำขอรวมทีมกับผู้ที่เข้าร่วมแข่งขันด้วยได้
    d. จำนวนผลที่ส่งได้ต่อวัน จะขึ้นอยู่กับจำนวนสมาชิกในทีม

87. Parameter ในข้อใดที่แสดงถึงคุณลักษณะภายในของกล้องที่ใช้ถ่ายภาพ
    a. Extrinsic parameter
    b. Transform parameter
    c. Projector parameter
    d. Intrinsic parameter

88. Crypto Wallets หาก MetaMask ปิดตัวไป เรายังสามารถเข้าถึง Wallet ของเราได้หรือไม่ ?
    a. สามารถเข้าถึงได้ โดยติดต่อทีมสนับสนุน MetaMask
    b. ไม่สามารถเข้าถึงได้
    c. สามารถเข้าถึงได้ โดยใช้ Seed Phrase บนกระเป๋าเงินอื่น
    d. สามารถเข้าถึงได้ โดยใช้บัญชีของ MetaMask บนเว็บไซต์ผู้ให้บริการอื่น

89. OpenCV สามารถเปิดวีดีโอจากแหล่งใดได้บ้าง
    a. ไฟล์
    b. IP camera
    c. ถูกทุกข้อ
    d. Webcam

90. ข้อใดเกี่ยวข้องกับกระบวนการ Information Extraction น้อยที่สุด
    a. Named Entity Recognition (NER)
    b. Part-of-Speech Tagging (PoS Tagging)
    c. Relation Extraction
    d. Image classification

91. Transfer Learning แบ่งการเรียนรู้เป็น 2 ครั้ง คือ Pre-traingและ Post-training
    a. TRUE
    b. FALSE