**1. ให้นศ. อธิบายคุณลักษณะของแต่ละ feature ในชุดข้อมูลที่ให้ ได้แก่ ความหมาย ช่วงค่าที่เป็นไปได้ ค่า outlier (ค่าที่ควรจะเป็นค่าผิดปกตินอกกรอบ)**

- uts: เป็นค่าวันและเวลาที่ทำการเก็บ sample นั้นๆ โดยมีช่วงอยู่ระหว่างวันที่ 2018-11-18 เวลา 08:18:41 ถึงวันที่ 2018-11-18 เวลา 16:09:04 ตามเวลาประเทศไทย ซึ่งพบว่าช่วงเวลาห่างไม่เท่ากัน

- accelerateX : เป็นค่าเร่งความเร็วในแกน X โดยมีช่วงอยู่ระหว่าง -19.497740 ถึง 9.685542 ค่าเฉลี่ย 2.227012 ไม่มีค่า outlier

- accelerateY : เป็นค่าเร่งความเร็วในแกน Y โดยมีช่วงอยู่ระหว่าง -22.248308 ถึง 7.831118 ค่าเฉลี่ย -3.975057 ไม่มีค่า outlier

- accelerateZ : เป็นค่าเร่งความเร็วในแกน Z โดยมีช่วงอยู่ระหว่าง -9.979478 ถึง 14.871238 ค่าเฉลี่ย 4.087197 ไม่มีค่า outlier

- compass: เป็นค่าที่บอกทิศทางของเข็มทิศ โดยมีช่วงอยู่ระหว่าง 35.128723 ถึง 800.000000 ค่าเฉลี่ย 81.645609 โดยมีค่า outlier เมื่อข้อมูลมีค่า > 400

- gps.x : เป็นค่าที่บอกพิกัดแกน x โดยมีช่วงอยู่ระหว่าง 0.000000 ถึง 13.621573 ค่าเฉลี่ย 1.457609 โดยมีค่า outlier เมื่อข้อมูลมีค่า == 0

- gps.y : เป็นค่าที่บอกพิกัดแกน x โดยมีช่วงอยู่ระหว่าง 0.000000 ถึง 100.369172 ค่าเฉลี่ย 10.772307 โดยมีค่า outlier เมื่อข้อมูลมีค่า == 0

- gyro.x : เป็นค่าที่บอกการหมุนในแกน x โดยมีช่วงอยู่ระหว่าง -3.799284 ถึง 4.511251 ค่าเฉลี่ย 0.001593 ไม่มีค่า outlier

- gyro.y : เป็นค่าที่บอกการหมุนในแกน y โดยมีช่วงอยู่ระหว่าง -5.185969 ถึง 4.016386 ค่าเฉลี่ย -0.009276 ไม่มีค่า outlier

- gyro.z : เป็นค่าที่บอกการหมุนในแกน z โดยมีช่วงอยู่ระหว่าง -3.459796 ถึง 3.534292 ค่าเฉลี่ย 0.027037 ไม่มีค่า outlier

- heartrate : เป็นค่าที่บอกอัตราการเต้นของหัวใจ โดยมีช่วงอยู่ระหว่าง 0.000000 ถึง 1000.000000 ค่าเฉลี่ย 83.344444 โดยมีค่า outlier เมื่อข้อมูลมีค่า == 0 และเมื่อข้อมูลมีค่า > 150

- light : เป็นค่าที่บอกความเข้มของแสง โดยมีช่วงอยู่ระหว่าง 0.000000 ถึง 2254.000000 ค่าเฉลี่ย 251.881772 โดยมีค่า outlier เมื่อข้อมูลมีค่า > 1000

- pressure : เป็นค่าที่บอกความดัน โดยมีช่วงอยู่ระหว่าง 0.000000 ถึง 1013.318000 ค่าเฉลี่ย 1007.298323 ไม่มีค่า outlier

**2. อธิบายการกำหนดค่า Outlier ที่ใช้และรูปแบบการจัดการสำหรับแต่ละ feature พร้อมเหตุผล**

- gps.x มีค่า outlier เมื่อข้อมูลมีค่า == 0 โดยเปลี่ยนในค่า outlier เป็นค่าก่อนของ row ก่อนหน้า เนื่องจากสันนิษฐานว่าตำแหน่งน่าจะอยู่ใกล้เคียงกันเมื่อเวลาใกล้กัน

- gps.y : มีค่า outlier เมื่อข้อมูลมีค่า == 0 โดยเปลี่ยนในค่า outlier เป็นค่าก่อนของ row ก่อนหน้า เนื่องจากสันนิษฐานว่าตำแหน่งน่าจะอยู่ใกล้เคียงกันเมื่อเวลาใกล้กัน

- heartrate : มีค่า outlier เมื่อข้อมูลมีค่า == 0 และเมื่อข้อมูลมีค่า > 150 เนื่องจากอัตราการเต้นของหัวใจถ้ามีค่าเป็น 0 คือเสียชีวิต ซึ่งในที่นี้ไม่น่าใช่ หรืออัตราการเต้นของหัวใจที่มากกว่า 150 อาจผิดปกติได้ จึงแทน outlier ด้วยค่าเฉลี่ยของ column heartrate

- light : มีค่า outlier เมื่อข้อมูลมีค่า > 1000 เนื่องจากข้อมูล มีค่ามากว่า 1000 แค่เพียงข้อมูลเดียว จึงแทน outlier ด้วยค่าเฉลี่ยของ column light

**3. ผลกระทบของ Feature Scaling ด้วยการใช้เทคนิคการ Norm ข้อมูล MaxMin Norm เทียบกับ zero mean & unit variance (x-mean/std)**

- MaxMin Norm : จะทำการ rescale ให้อยู่ในช่วง [0,1] โดยจำเป็นต้องจัดการเรื่องค่า outlier ก่อนการ rescale

- zero mean & unit variance (x-mean/std): จะทำการ rescale ให้ค่าเฉลี่ยมีค่าเท่ากับ 0 และการกระจายมีค่าเท่ากับ 1 โดยอยู่ในช่วงระหว่าง [-3, +3]

**4.ในชุด Feature ที่ให้ Feature ใด เหมาะกับการทำ Feature Scaling แบบ Max-min Norm, Standard Norm, หรือ ไม่สามารถทำ Feature scaling ได้**

- Max-min Norm : accelerateX, accelerateY, accelerateZ , gyro.x, gyro.y และ gyro.z

- Standard Norm : compass, heartrate, light และpressure

- ไม่สามารถทำ Feature scaling : gps.x และgps.y

**5.การทำ Moving average ด้วย window size ที่แตกต่างกัน ให้ผลต่างกันอย่างไร**

- เมื่อ window size แตกต่างกัน ให้ผลต่างกันคือ เมื่อ window size มีขนาดเล็ก จะทำให้กราฟที่พลอต ไม่อาจทำให้เห็นหลักเทรนของข้อมูล แต่ถ้า window size มีขนาดใหญ่อาจจะทำให้กราฟที่พลอตพลาดรายละเอียดที่สำคัญที่สนใจไปได้

**6.การทำ Data Slicing หรือการจัดชุดข้อมูล เพื่อนำมาเตรียมเป็นข้อมูลใช้กับโมเดล ให้นศ. ระบุ shape ของ A,B พร้อมขั้นตอนในการ ทำให้ผลลัพธ์ เปลี่ยนจาก A -> B ได้ตามรูปที่กำหนด**

- A.shape = (3, 3)

- B.shape = (3, 3, 1)

- A.reshape(3,3,1)

**7.ค่า PCA ที่ได้จากการทดลองตอนที 2 นศ.คิดว่า น้ำหนักการผสม feature input ค่าใดสำคัญที่สุดในชุดข้อมูลนี้ เพราะเหตุใด**

- accelerateY และ accelerateZ เนื่องจากมีค่า eigenvalue มากที่สุด 2 อันดับแรก

**8.นศ.คิดว่า PCA ของข้อมูลในการทดลอง ถ้าต้องลด dimension ของข้อมูล จาก 3D -> 2D สามารถทำได้อย่างมั่นใจ หรือควรนำข้อมูลทั้ง 3D และ 2D ไป ทดสอบกับ model ก่อนตัดสินใจลดข้อมูล เพราะเหตุใด**

- สามารถดูที่ค่า variation ของข้อมูลที่ถูกแปลงด้วย PCA โดยจะต้องคลอบคลุม 80% ของข้อมูลทั้งหมด