การรู้จำอิริยาบถของมนุษย์โดยใช้ข้อมูลจากเซนเซอร์ของสมาร์ทโฟน สำหรับการยืนยันตัวตนผู้ใช้งาน

Human Activity Recognition using Sensor Data of Smartphones for User Authentication

นางสาวณัฐพร โสมหิรัญ

Natthaporn Somhirun

นางสาวรุ่งเรืองทรัพย์ สุขปรุง

Rungruangsap Sukprung

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ ปีการศึกษา พ.ศ. 2561

คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

โครงงาน	การรู้จำอิริยาบถของมนุษย์โดยใช้ข้อมูลจากเซนเซอร์ของสมาร์ทโฟนสำหรับการยืนยันตัวตนผู้ใช้งาน Human Activity Recognition using Sensor Data of Smartphones for User Authentication	
นิสิต	นางสาวณัฐพร โสมหิรัญ	58102010804
	นางสาวรุ่งเรืองทรัพย์ สุขปรุง	58102010822
ปริญญา	**วิทยาศาสตรบัณฑิต(วท.บ.)**	
ภาควิชา	**วิทยาการคอมพิวเตอร์**	
อาจารย์ที่ปรึกษาโครงงาน	**ผู้ช่วยศาสตราจารย์ ดร.วราภรณ์ วิยานนท์**	

ลงชื่อ.....
(ผู้ช่วยศาสตราจารย์ ดร.วราภรณ์ วิยานนท์)

บทคัดย่อ

สมาร์ทโฟนเป็นอุปกรณ์ที่ช่วยผู้ใช้ในการอำนวยความสะดวกในการทำงานต่างๆ ได้ทุกที่ทุกเวลา เช่น การรับส่งอีเมล การทำธุรกรรมอิเล็กทรอนิกส์และการบันทึกข้อมูลต่างๆ ผ่านสมาร์ทโฟน โดยผู้ใช้จำนวนมาก ใช้สมาร์ทโฟนในการจัดเก็บข้อมูล เช่น ข้อมูลส่วนบุคคลและรายละเอียดบัญชีธนาคารซึ่งข้อมูลที่มีความสำคัญ เหล่านี้อาจตกอยู่ในความเสี่ยงถ้าหากอุปกรณ์ถูกขโมยหรือสูญหาย วิธีการดั้งเดิมในการปกป้องข้อมูลประเภทนี้บนสมาร์ทโฟนคือการตรวจสอบผู้ใช้ด้วยวิธีต่างๆ เช่น PIN รหัสผ่านและการจดจำลายนิ้วมือ ซึ่งวิธีเหล่านี้เป็น การยืนยันตัวตนแบบแอคทีฟที่อาจมีความเสี่ยงต่อการถูกโจมตี วิธีการยืนยันตัวตนที่จะทำให้มีความปลอดภัย ต่อข้อมูลภายในสมาร์ทโฟนมากขึ้นคือการยืนยันตัวตนแบบพาสซีฟและต่อเนื่อง คือ ผู้ใช้งานไม่จำเป็นต้องทำ การยืนยันตัวตนด้วยตนเองโดยตรงและทำการยืนยันซ้ำในช่วงระยะเวลาหนึ่ง

ดังนั้น เราจึงได้พัฒนาแอปพลิเคชันที่มีการยืนยันตัวตนแบบพาสซีฟและต่อเนื่อง เพื่อลดความเสี่ยงต่อ การถูกโจมตีจากความพยายามในการเข้าใช้งานสมาร์ทโฟน โดยใช้ลักษณะทางพฤติกรรม (behavioral biometrics) เช่น การยืน การเดินและการวิ่ง เพื่อตรวจสอบผู้ใช้งานสมาร์ทโฟนอย่างต่อเนื่อง ในปัจจุบัน สมาร์ทโฟนมีเซนเซอร์ที่ให้ข้อมูลเพียงพอต่อการรู้จำอิริยาบถของมนุษย์ซึ่งสามารถนำข้อมูลเหล่านี้มาวิเคราะห์ ด้วยวิธีการเรียนรู้ของเครื่อง (machine learning) ได้ โดยเราได้ทำการพัฒนาโมเดลขึ้นมา 2 โมเดลสำหรับ การพัฒนาแอปพลิเคชันคือโมเดลสำหรับการรู้จำอิริยาบถและโมเดลสำหรับการยืนยันตัวตนผู้ใช้งาน สมาร์ทโฟน โดยใช้ข้อมูลจากเซนเซอร์สมาร์ทโฟนที่เก็บจากการวางตำแหน่งสมาร์ทโฟน 4 ตำแหน่งที่ตัวผู้ใช้ งานสมาร์ทโฟน ได้แก่ กระเป๋ากางเกงข้างขวา ข้อมือ ต้นแขน และเข็มขัด โมเดลที่ 1 : โมเดลสำหรับการ รู้จำอิริยาบถ มี 3 ขั้นตอนหลัก ได้แก่ (1) การเตรียมข้อมูล (Data pre-processing) (2) การสกัดคุณลักษณะ (Feature extraction) (3) นำข้อมูลเข้าสู่กระบวนการฝึกสอนโดยใช้ SVM (Support Vector Machines), kNN (k-Nearest Neighbor), Decision Tree และ Bayesian networks เป็นอัลกอริทึมการเรียนรู้ของ เครื่องที่ใช้ในการจำแนกอิริยาบถ โมเดลที่ 2 : โมเดลสำหรับการยืนยันตัวตนผู้ใช้งานสมาร์ทโฟน มี 3 ขั้นตอนหลัก ได้แก่ (1) การเตรียมข้อมูล (Data pre-processing) (2) การสกัดคุณลักษณะ (Feature extraction) (3) นำข้อมูลเข้าสู่กระบวนการฝึกสอนโดยใช้ One-class SVM และ One-class kNN ในการยืนยันตัวตนผู้ใช้งาน

จากผลการทดสอบการรู้จำอิริยาบถ โมเดลจากการการฝึกสอนโดยใช้ SVM (Support Vector Machines) มีประสิทธิภาพการรู้จำอิริยาบถที่ดีที่สุดด้วยประสิทธิภาพ 99.48% บริเวณตำแหน่งต้นแขน ส่วนการยืนยันตัวตนผู้ใช้งานสมาร์ทโฟน โมเดลจากการฝึกสอนโดยใช้ one-class SVM ให้ค่าที่ดีที่สุด ในตำแหน่งเข็มขัด อิริยาบถการนั่ง 97.35% โดยโมเดลที่ได้จะนำไปพัฒนาแอปพลิเคชันการยืนยันตัวตน แบบพาสซีฟและต่อเนื่องต่อไป

Abstract

Smartphones are devices that help users to facilitate various functions at anywhere in anytime such as sending and receiving an email, electronic transactions, and their private data. Many users use a smartphone to store personal information and bank account details which are sensitive information. These may be at risk if the device is stolen or lost. To protect users' data on mobile devices in traditional ways the user authentication approaches are PINs, passwords, and fingerprint recognition which are active user authentication. Passive and continuous user authentication will secure users' data in the smartphones that the user does not need to manually verify their identity and repeat the verification for a certain period of time.

Therefore, we have developed an application that has passive and continuous user authentication to reduce the risk of being attacked by any attempts to access the smartphone by using behavioral biometrics such as standing, walking and running to continuously monitor smartphone users. Nowadays, smartphones have sensors that provide enough information to recognize human activity which we can use this information to analyze human activity using machine learning methods. We developed the models for the application development which can be classified into two segments: 1) models for activity recognition and 2) models for user authentication. The data from smartphone sensors collected from four positions of the user's body, including the right pocket, wrist, upper arm, and belt. To develop the models for activity recognition, there are three steps consisting of (1) data pre-processing (2) feature extraction (3) training process using SVM, k-NN, Decision tree, and Bayesian networks. To develop the models for user authentication there are four steps consisting of (1) data pre-processing (2) feature extraction (3) feature selection (4) training process using one-class SVM and one-class k-NN to verify user authentication.

From experimental results for activity recognition, the model from SVM can achieve excellent activity recognition performance with 99.48% on an upper arm. For the user authentication model, one-class SVM returns the magnificent results of 97.35% with sitting activity and belt position. These models will be used to develop a passive and continuous user authentication application.

กิตติกรรมประกาศ

โครงงานการรู้จำอิริยาบถของมนุษย์โดยใช้ข้อมูลจากเซนเซอร์ของสมาร์ทโฟนสำหรับการยืนยันตัวตนผู้ใช้งาน Human Activity Recognition using Sensor Data of Smartphones for User Authentication สามารถดำเนินการได้สำเร็จลุล่วง โดยได้รับความช่วยเหลือ และความอนุเคราะห์จากบุคคล ดังต่อไปนี้

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ดร. วราภรณ์ วิยานนท์ อาจารย์ที่ปรึกษาโครงการผู้ให้ความรู้ และคำแนะนำเกี่ยวกับแหล่งค้นคว้าศึกษาข้อมูลที่เป็นประโยชน์ รวมถึงช่วยแนะนำแนวทางในการแก้ไขปัญหา ต่างๆที่เกิดขึ้นระหว่างการจัดทำโครงการตลอดจนการตรวจสอบแก้ไขข้อบกพร่องในการทำโครงการ

ขอขอบพระคุณอาจารย์ทุกท่านในภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่กรุณาให้คำแนะนำข้อคิดเห็นและความรู้ต่างๆที่เป็นประโยชน์ต่อการจัดทำโครงการวิจัย

สุดท้ายนี้ขอขอบพระคุณบิดามารดาและครอบครัว ที่เปิดโอกาสให้ศึกษาเล่าเรียนรวมถึงให้ความช่วยเหลือและให้กำลังใจผู้จัดทำจนโครงการวิจัยประสบความสำเร็จเป็นอย่างดี

สารบัญ

TOC \o "1-3" \h \z \u **บทคัดย่อ** 3

Abstract 4

กิตติกรรมประกาศ 5

สารบัญ 6

สารบัญภาพ 8

สารบัญตาราง 9

บทที่ 1 บทนำ 10

1.1 ที่มาและความสำคัญ 10

1.2 วัตถุประสงค์ของงานวิจัย 10

1.3 ขอบเขตของการวิจัย 10

1.4 ประโยชน์ที่คาดว่าจะได้รับ 11

บทที่ 2 องค์ความรู้ที่เกี่ยวข้อง 12

2.1 ทฤษฎีและแนวคิดที่ใช้ในการทำวิจัย 12

2.1.1 Machine learning 12

2.1.2 Data Pre-Processing 14

2.1.3 การประยุกต์ใช้การทำเหมืองข้อมูล (Data Mining) เพื่อรู้จำอิริยาบถ 15

2.1.4 การยืนยันตัวตนผู้ใช้งานจากการรู้จำอิริยาบถ 16

2.1.5 การประเมินประสิทธิภาพ (Evaluation) 18

2.2 งานวิจัยที่เกี่ยวข้อง 19

2.2.1 Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme [3] 19

2.2.2 Multi-sensor authentication to improve smartphone security [4] 19

2.2.3 Smartphone-Based Gait Recognition From Authentication to Imitation [5] 20

บทที่ 3 วิธีการดำเนินงานศึกษาค้นคว้า 21

3.1 วิธีการดำเนินการวิจัย 21

3.2 แผนการดำเนินงานตลอดโครงงาน 21

3.3 อุปกรณ์และเครื่องมือที่ใช้ในการวิจัย 23
3.3.1 ฮาร์ดแวร์ 23
3.3.2 ซอฟต์แวร์ 23
3.3.2 ภาษาที่ใช้ 23
3.4 การออกแบบและพัฒนา 23
3.4.1 ชุดข้อมูล (Dataset) 24
3.4.2 การรู้จำอิริยาบถ (Activity Recognition) 26
3.4.3 การยืนยันตัวตนผู้ใช้งานจากการรู้จำอิริยาบถ (User authentication for activity recognition) 29
3.4.4 การเก็บข้อมูลเพื่อใช้สำหรับยืนยันตัวตนผู้ใช้งาน 32
3.4.5 พัฒนาแอปพลิเคชันสำหรับยืนยันตัวตนผู้ใช้งาน 34
บทที่ 4 ผลการดำเนินงาน 35
4.1 ผลการประเมินประสิทธิภาพของการรู้จำอิริยาบถ 35
4.2 ผลการประเมินประสิทธิภาพของการยืนยันตัวตน 36
บทที่ 5 สรุปผลการดำเนินงาน 40
5.1 สรุปผลการดำเนินงาน 40
5.2 ปัญหาและอุปสรรคในการดำเนินงาน 41
5.3 ข้อเสนอแนะ 41
บรรณานุกรม 42
ภาคผนวก 43
ภาคผนวก ก 44
ภาคผนวก ข 50
ภาคผนวก ค 55

สารบัญภาพ

TOC \h \z \c "ภาพที่" ภาพที่ 1 รูปแบบการเรียนรู้ของ Machine Learning 12
ภาพที่ 2 รูปแบบการเรียนรู้แบบมีผู้สอน 13

สารบัญตาราง

TOC \h \z \c "ตารางที่" ตารางที่ 1 แผนการดำเนินงาน 21

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญ

สมาร์ทโฟนเป็นอุปกรณ์ที่ช่วยผู้ใช้ในการอำนวยความสะดวกในการทำงานต่างๆ ได้ทุกที่ทุกเวลา เช่น การรับส่งอีเมล การทำธุรกรรมอิเล็กทรอนิกส์ผ่านสมาร์ทโฟน ลักษณะของงานที่ดำเนินการกับอุปกรณ์เหล่านี้ ได้พัฒนาขึ้นเรื่อยๆ เนื่องจากผู้ใช้จำนวนมากใช้สมาร์ทโฟนในการจัดเก็บข้อมูลส่วนตัว เช่น ข้อมูลส่วนบุคคล และรายละเอียดบัญชีธนาคาร อย่างไรก็ตามข้อมูลสำคัญเหล่านี้อาจมีความเสี่ยงหากอุปกรณ์ ถูกขโมย หรือสูญหายวิธีการดั้งเดิมในการปกป้องข้อมูลประเภทนี้บนสมาร์ทโฟนคือการตรวจสอบผู้ใช้ด้วยกลไก ต่างๆเช่นPINรหัสผ่านและการจดจำลายนิ้วมือซึ่งอาจมีความเสี่ยงต่อการถูกโจมตีเนื่องจากการยืนยันตัวตน โดยวิธีการข้างต้นไม่สามารถทำการยืนยันตัวตนแบบพาสซีฟและต่อเนื่องได้

การยืนยันตัวตนแบบพาสซีฟและต่อเนื่องมีจุดมุ่งหมายเพื่อแก้ไขปัญหาเหล่านี้โดยเลือกใช้ลักษณะทาง พฤติกรรม (behavioral biometrics) เช่นการเดินเพื่อตรวจสอบผู้ใช้สมาร์ทโฟนอย่างต่อเนื่อง และในปัจจุบัน สมาร์ทโฟนมีเซนเซอร์ที่ให้ข้อมูลเพียงพอต่อการรู้จำอิริยาบถของมนุษย์ข้อมูลเหล่านี้ สามารถนำมาวิเคราะห์ ด้วยวิธีการเรียนรู้ของเครื่อง (machine learning) เพื่อยืนยันตัวตนในการเข้าถึงข้อมูลภายในสมาร์ทโฟน

ด้วยเหตุนี้ ผู้วิจัยเล็งเห็นว่าการยืนยันตัวตนแบบพาสซีฟและต่อเนื่องโดยใช้การรู้จำอิริยาบถของผู้ใช้ จะทำให้มีความปลอดภัยมากขึ้นต่อข้อมูลภายในสมาร์ทโฟน

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาและพัฒนาวิธีการรู้จำอิริยาบถโดยใช้ข้อมูลจากเซนเซอร์ของสมาร์ทโฟน
2. เพื่อศึกษาและพัฒนาวิธีการยืนยันตัวตนผู้ใช้งานโดยใช้ข้อมูลจากเซนเซอร์ของสมาร์ทโฟน
3. เพื่อสร้างแอปพลิเคชันสำหรับการยืนยันตัวตน โดยใช้การรู้จำอิริยาบถของผู้ใช้งาน

1.3 ขอบเขตของการวิจัย

1. ใช้ชุดข้อมูล 2 ชุด จาก SWELL project ในเว็บไซต์มหาวิทยาลัย University of Twente [1]
    - Physical activity recognition dataset สำหรับการสร้างโมเดลการรู้จำอิริยาบถ
    - Sensors activity dataset สำหรับการทดสอบการยืนยันตัวตน
2. ใช้ Classification model ในการจำแนกอิริยาบถและยืนยันตัวตนของผู้ใช้งานสมาร์ทโฟน
3. ใช้ Python ในการสร้างโมเดล
4. พัฒนาแอปพลิเคชันต้นแบบสำหรับการยืนยันตัวตนผู้ใช้งาน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เรียนรู้ Machine learning
2. ได้ทดสอบการใช้เทคนิคการระบุประเภทข้อมูลของ Machine learning ในการรู้จำอิริยาบถของมนุษย์
3. เรียนรู้วิธีการยืนยันตัวตนผู้ใช้งานสมาร์ทโฟนจากการรู้จำอิริยาบถ
4. แอปพลิเคชันสำหรับการยืนยันตัวตนโดยใช้การรู้จำอิริยาบถ

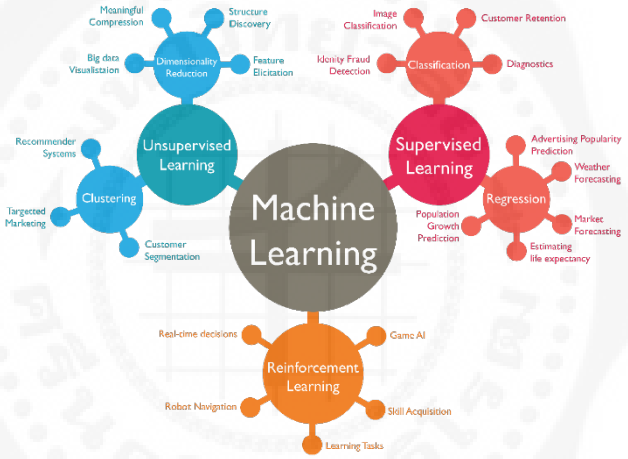
บทที่ 2
องค์ความรู้ที่เกี่ยวข้อง

2.1 ทฤษฎีและแนวคิดที่ใช้ในการทำวิจัย

2.1.1 Machine learning

Machine learning เป็นศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ (Artificial inteligence) ที่ทำให้คอมพิวเตอร์มีความสามารถในการเรียนรู้ด้วยตนเองโดยใช้ข้อมูล ซึ่งจะใส่ข้อมูล (Data) และผลลัพธ์ (Output) เข้าไปเพื่อให้ได้คำตอบว่าข้อมูลนำเข้า (Input) ที่นำเข้าไปจะได้ผลลัพธ์เป็นอะไร ซึ่งแตกต่างกับการเขียนโปรแกรมทั่วไปที่ใส่ข้อมูลเข้าไปเพื่อให้ได้ผลลัพธ์การใช้ข้อมูลนั้น สามารถใช้ได้หลาย แบบ ซึ่งจะแบ่งตามรูปแบบการเรียนรู้ของ Machine Learning

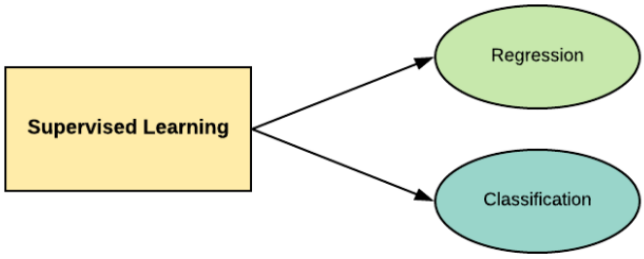
รูปแบบการเรียนรู้ของ Machine learning มี 3 รูปแบบดังภาพที่ 1 ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised learning), การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) และการเรียนรู้แบบเสริม กำลัง (Reinforcement learning)



ภาพที่ 1 รูปแบบการเรียนรู้ของ Machine Learning
ที่มา : https://blog.finnomena.com/machine-learning-คืออะไร-fa8bf6663c07

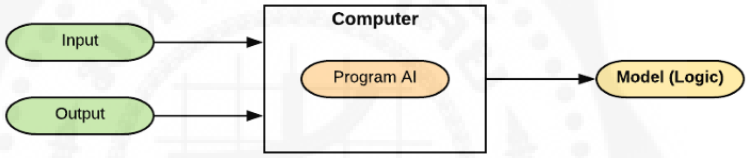
1) การเรียนรู้แบบมีผู้สอน (Supervised learning)

การเรียนรู้แบบมีผู้สอนเป็นการเรียนรู้โดยอาศัยข้อมูลและเลเบล (label) จากข้อมูลสำหรับ การเรียนรู้ โดยสามารถแบ่งการเรียนรู้แบบมีผู้สอนได้ 2 แบบดังภาพที่ 2 คือ 1)การแบ่งประเภทข้อมูล(classification) โดยผลลัพธ์จะมีลักษณะเป็นค่าไม่ต่อเนื่อง (Discretevalue) คือค่าที่บอก เป็นกลุ่มของข้อมูลซึ่งจะสามารถกำหนดไว้ล่วงหน้าได้และ 2) การวิเคราะห์การถดถอย (regression) ได้ผลลัพธ์เป็นค่าต่อเนื่อง(Continuousvalues) คือ ผลลัพธ์ที่ได้จะออกมาเป็นค่าตัวเลขซึ่งเหมาะแก่ การใช้ในการทำนาย

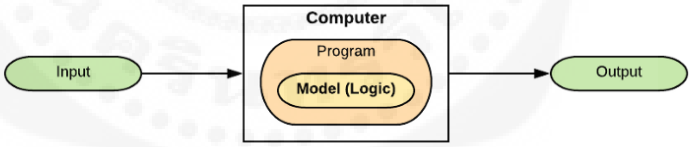


ภาพที่ 2 รูปแบบการเรียนรู้แบบมีผู้สอน
ที่มา : https://medium.com/@every.phu/supervised-learning--คืออะไร-ทำงานยังไง-1c0e411a

การเรียนรู้แบบมีผู้สอนมีรูปแบบการเขียนโปรแกรมที่แตกต่างออกไป โดยให้คอมพิวเตอร์ สร้างโมเดลขึ้นมาจากข้อมูลนำเข้าและผลลัพธ์ที่ต้องการดังภาพที่ 3 จากนั้นจึงนำโมเดลที่ได้มา ประยุกต์ใช้ดังภาพที่ 4 ดังนั้นหากมีข้อมูลนำเข้าและผลลัพธ์ที่มีความหลากหลายและจำนวน มากจะทำให้มีโอกาสได้โมเดลที่มีความแม่นยำมากขึ้น โดยกระบวนการสร้างโมเดลแบบนี้เรียกว่า การเทรน (training)



ภาพที่ 3 กระบวนการเทรน เพื่อให้ได้ model ที่ต้องการ
ที่มา : https://medium.com/@every.phu/supervised-learning--คืออะไร-ทำงานยังไง-1c0e411a40a2



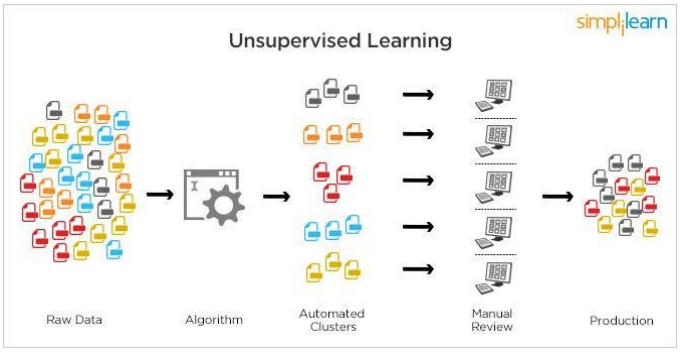
ภาพที่ 4 กระบวนการนำโมเดลไปประยุกต์ใช้กับโปรแกรม

ที่มา : https://medium.com/@every.phu/supervised-learning--คืออะไร-ทำงานยังไง-1c0e411a40a2

2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning)

การเรียนรู้แบบไม่มีผู้สอนเป็นการเรียนรู้ที่ไม่จำเป็นต้องกำหนดเลเบล (label) บอกว่าข้อ มูลนั้นคืออะไร โดยจากภาพที่ 5 การเรียนรู้นี้คือ การที่นำข้อมูล (Input) เข้าแต่ไม่ได้ระบุว่าข้อมูล ที่นำเข้านั้นเป็นอะไรหลังจากนั้นจะนำเอาไปวิเคราะห์ (Feature extraction) ว่าข้อมูลที่นำเข้าไปมี ลักษณะอย่างไร แต่ไม่สามารถเอาไปแบ่งประเภทข้อมูล (Classification) ได้ จึงใช้

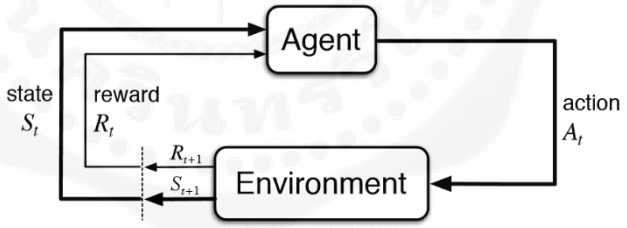
วิธีการจัดเป็นกลุ่ม(cluster) บนพื้นฐานของความเหมือน(similarities)และความแตกต่าง (differences) ของรูปแบบข้อมูลนำเข้า



ภาพที่ 5 กระบวนการการเรียนรู้แบบไม่มีผู้สอน
ที่มา : https://i.pinimg.com/736x/1e/3a/4e/1e3a4ed9359b5efc8d227ecae095d7a2--machine-learning-big-data.jpg

3) การเรียนรู้แบบเสริมกำลัง (Rainforcement learning)

เป็นเทคนิคการเรียนรู้ที่คอมพิวเตอร์มีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนไปตลอดเวลาโดยคอมพิวเตอร์จะต้องทำงานบางอย่างโดยที่ไม่รู้วิธีการที่ทำอยู่นั้นเข้าใกล้เป้าหมายแล้วหรือไม่ ดังภาพที่ 6 ตัวอย่างเช่น การเล่นหมากรุก จะต้องมีการทำนายล่วงหน้าว่าจะสามารถเกิดอะไรขึ้นได้ ซึ่งการเดินแต่ละครั้งอาจจะไม่เป็นผลดีต่อครั้งนั้นแต่อาจมีผลดีในครั้งต่อไปได้



ภาพที่ 6 ภาพแสดงกระบวนการการเรียนรู้แบบเสริมกำลัง
ที่มา : https://medium.com/kosate/dqn-deep-learning

2.1.2 Data Pre-Processing

1. Noise Removal ข้อมูลที่บันทึกจากเซ็นเซอร์ของสมาร์ทโฟนมีสัญญาณรบกวนที่เกิดจากผู้ใช้และตัวเซ็นเซอร์เอง โดยจำเป็นในการกำจัดสัญญาณรบกวนจากข้อมูลก่อนการประมวลผล
    - Average smoothing method การแทนที่ข้อมูลดิบจากเซนเซอร์ด้วยค่าเฉลี่ยของจุดข้อมูลที่ติดกัน

$$y[i] = \frac{1}{M}\sum_{j=0}^{M-1} y[i+j]$$

โดยที่ y[i] คือ ค่าเฉลี่ยของข้อมูลตำแหน่งที่ i

M คือ จำนวนของจุดข้อมูลที่ใช้ในการหาค่าเฉลี่ย

2. Data Segmentation
   - การเพิ่มค่า magnitude คำนวนณจากทั้ง 3 แกนของแต่ละเซนเซอร์ดังสมการที่กำหนด

$$magnitude = \sqrt{x^2 + y^2 + z^2}$$

เมื่อ x, y และ z แทนค่า x-axis, y-axis และ z-axis ตามลำดับ

   - เพื่อให้สามารถ feature extraction และฝึกสอนการจำแนกได้ดี จำเป็นต้องแบ่งข้อมูลเซ็นเซอร์ตามแต่ละแกนออกเป็นส่วนย่อย ๆ โดยใช้วิธี fixed-size window segmentation เพื่อจุดประสงค์นี้ [2]

2.1.3 การประยุกต์ใช้การทำเหมืองข้อมูล (Data Mining) เพื่อรู้จำอิริยาบถ

Data Mining คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงาน หลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหารในด้านวิทยาศาสตร์และการแพทย์ รวมทั้งในด้านเศรษฐกิจและสังคม

การรู้จำอิริยาบถด้วยแนวทางการทำเหมืองข้อมูล (Data mining base approach to activity recognition) ซึ่งได้นำข้อมูลที่ได้ จากการเดิน การนั่ง การยืน การวิ่ง การเดินขึ้นบันได และการเดินลงบันได นำมาวิเคราะห์ด้วยวิธีการจำแนก (Classification) โดยใช้ Python สร้างโมเดลเพื่อจำแนกข้อมูลการเคลื่อนไหวที่ตรวจจับด้วยเซนเซอร์ในสมาร์ทโฟน ผลลัพธ์ที่ได้คือ เกณฑ์การจำแนกอิริยาบถต่างๆ ได้แก่ การเดิน การนั่ง การยืน การวิ่ง การเดินขึ้นบันได และการเดินลงบันได อัลกอริทึมที่ใช้ในการจำแนก ได้แก่

1) k-Nearest Neighbors

k-Nearest Neighbor (kNN) เป็นวิธีการจำแนกประเภทข้อมูล (Data Classification) วิธีการหนึ่ง โดยจัดเป็นวิธีการจำแนกประเภทข้อมูลแบบมีผู้ฝึกสอน (Supervised Learning) หรือการที่ทราบคำตอบของข้อมูลอยู่ก่อนแล้วจากนั้นใช้โมเดลในการจำแนกประเภทข้อมูลจากข้อมูลฝึกสอนที่ทราบคำตอบวิธีการจำแนกจะใช้วิธีการวิเคราะห์จาก

ข้อมูลที่ใกล้เคียงที่สุดจำนวน k ตัว กับข้อมูลที่ต้องการจำแนกประเภทของข้อมูลหรือต้องการทำนายคลาสของข้อมูลใหม่โดยจะทำนาย ตามคลาสส่วนใหญ่ของข้อมูลฝึก k ตัว

2) Support Vector Machines

Support Vector Machines เป็นอัลกอริทึมที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูล ใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการ เพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยัง เส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด

3) Decision Tree

Decision Tree เป็นการเรียนรู้โดยการจำแนกประเภท (classification) ข้อมูลออกเป็นกลุ่ม (class) ต่างๆ โดยใช้คุณลักษณะ (attribute) ข้อมูลในการจำแนกประเภท ทำให้ทราบว่า คุณลักษณะใดเป็นตัวกำหนดการจำแนกประเภทและคุณลักษณะแต่ละตัวมีความสำคัญมากน้อยต่างกันอย่างไร

4) Bayesian networks

การจำแนกข้อมูลด้วยวิธีการเบย์เซียนจะเป็นการสร้างตัวจำแนกข้อมูลด้วยการประยุกต์ใช้ค่าทางสถิติ ซึ่งจะสามารถบ่งบอกถึงความน่าจะเป็นของข้อมูลเรคคอร์ดหนึ่งที่จะอยู่ในหมวดหมู่ของ ข้อมูลหนึ่งๆวิธีการจำแนกข้อมูลนี้จะทำการประยุกต์ใช้ทฤษฎีของเบย์ (Bayes' theorem) ซึ่งจะช่วย ให้สามารถจำแนกข้อมูลได้อย่างรวดเร็ว

2.1.4 การยืนยันตัวตนผู้ใช้งานจากการรู้จำอิริยาบถ

การยืนยันตัวตนผู้ใช้งานจากการรู้จำอิริยาบถเป็นการยืนยันตัวตนเป็นการใช้ข้อมูลไบโอเมตริกที่ได้จากการสกัดคุณลักษณะของแต่ละอิริยาบถซึ่งเป็นคุณลักษณะที่เป็นเอกลักษณ์เฉพาะแต่ และผู้ใช้งาน ทำให้สามารถแยกแยะ หรือจดจำผู้ใช้งานได้

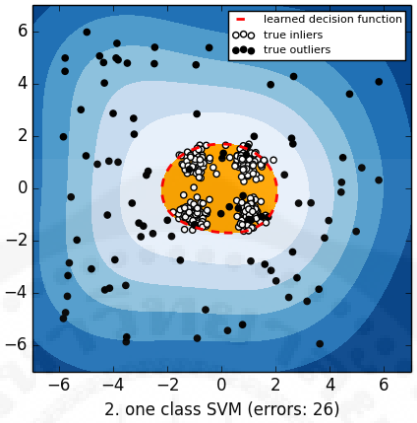
รูปแบบการใช้ข้อมูลไบโอเมตริกเพื่อการยืนยันตัวตน

1. การระบุตัวตนผู้ใช้งาน(Identification) หรือการจับคู่เปรียบเทียบแบบหนึ่งต่อ จำนวน มากกว่าหนึ่ง (1:N) การทำงานจะนำข้อมูลไบโอเมตริกที่ได้รับโดยที่ไม่มี การระบุตัวตนผู้ใช้งานไปทำการจับคู่เปรียบเทียบกับข้อมูลทั้งหมดในฐานข้อมูลเพื่อระบุว่าข้อมูลที่ได้รับเป็นข้อมูลของบุคคลใด
2. การตรวจพิสูจน์ตัวตนผู้ใช้งาน (Verification) หรือการจับคู่เปรียบเทียบแบบหนึ่งต่อ หนึ่ง (1:1) การทำงานจะนำข้อมูลไบโอเมตริกที่ ได้รับและมีการระบุตัวตนผู้ใช้งาน แล้วมาเปรียบเทียบดูว่าข้อมูลที่ได้รับมาตรงกับข้อมูลที่ได้ถูกบันทึกไว้ก่อนหน้านี้ หรือไม่ โดยทั่วไปการทำงานของกระบวนการนี้นั้นใช้เวลาไม่มาก เพราะเป็นการ เปรียบเทียบแบบหนึ่งต่อหนึ่ง จำนวนข้อมูลที่ต้องเปรียบเทียบไม่มากเหมือนอย่าง กรณีการระบุตัวตนผู้ใช้งาน

ในขั้นตอนการยืนยันตัวตนระบบจะตัดสินใจว่าผู้ใช้นั้นเป็นผู้ใช้ที่ถูกต้องหรือเป็นผู้แอบอ้าง โดยวิธีการทีเลือกใช้คือ

1) One-class SVM

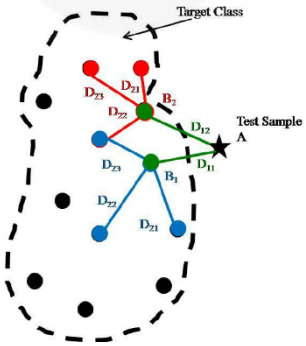
เป็นอัลกอริทึมที่ปรับวิธีการมาจาก SVM เพื่อแก้ปัญหาการจำแนกประเภทของคลาสเดียว คำนวณฟังก์ชันเลขฐานสอง ซึ่งจะสร้างทรงกลมไฮเปอร์รอบข้อมูลนำเข้าซึ่งครอบคลุมเกือบทุกจุด ในชุดข้อมูลด้วยรัศมีต่ำสุดหากข้อมูลใดอยู่นอกทรงกลมจะถูกนำออก โดยอาจมีข้อมูลบางส่วนที่ถูก นำออกไปเป็นข้อมูลที่ถูกต้องจึงต้องปรับทรงกลมยืดหยุ่นได้มากขึ้นโดยฟังก์ชั่นเคอร์เนล ตัวอย่างดังภาพที่ 7



ภาพที่ 7 ภาพแสดงการทำงานของ One-class SVM

2) One-Class k-Nearest Neighbor

วิธีนี้จะประมาณค่าความหนาแน่นโดยใช้ระยะทางของข้อมูลที่ใกล้ที่สุดตัวแรกเท่านั้น โดยมีเงื่อนไขว่าความหนาแน่นของข้อมูลนำเข้านั้นต้องมากกว่าหรือเท่ากับความหนาแน่นของข้อมูลชุดฝึกสอนการกำหนดค่า k นั้นจะแตกต่างกันไป โดยการเพิ่มค่า k จะทำให้การตรวจจับการจำแนกช้าแต่มีสัญญาณรบกวนน้อยลง โดยสามารถใช้ระยะทางเฉลี่ยของ k กับ k ตัวแรกเป็นเกณฑ์ของค่าได้ สามารถเปลี่ยนเป็นค่าที่สูงกว่าหรือต่ำกว่าเพื่อเปลี่ยนความไวในการตรวจจับการจำแนก ตัวอย่างดังภาพที่ 8



ภาพที่ 8 ภาพแสดงการทำงานของ One-class k-NN

โดย One-class k-NN สามารถตัดสินใจได้จาก

- ให้ข้อมูลนำเข้าเป็นค่า z โดยจะหา nearest neighbor ของ z จากชุดฝึกสอน โดยจะเรียกว่า y
- หลังจากนั้นจะทำการหาค่า nearest neighbor ของ y โดยกำหนดให้เป็น NN(y)
- จากสูตร จะจำแนกว่า z เป็นข้อมูลที่เป็นเป้าหมายหรือไม่

ำ Threshold คือ 1 แต่สามารถเลือกได้ตามความต้องการ

2.1.5 การประเมินประสิทธิภาพ (Evaluation)

การประเมินประสิทธิภาพของโมเดลสามารถทำได้โดยการทดสอบการจำแนกประเภทข้อมูลด้วยข้อมูลทดสอบ (Testing Data) และทำการประเมินประสิทธิภาพของโมเดล โดยโมเดลการรู้จำ อิริยาบถใช้ค่าความถูกต้อง (Accuracy) ซึ่งหมายถึง ค่าอัตราส่วนที่โมเดลสามารถจำแนกข้อมูลได้ อย่างถูกต้อง ซึ่งสามารถคำนวณได้จากสมการ

โดย

TP หรือ True Positive คือ จำนวนข้อมูลที่มีการจำแนกคลาสเป้าหมาย Positive เป็น Positive

TN หรือ True Negative คือ จำนวนข้อมูลที่มีการจำแนกคลาสเป้าหมาย Negative เป็น Negative

FP หรือ False Positive คือ จำนวนข้อมูลที่มีการจำแนกคลาสเป้าหมาย Negative เป็น Positive

FN หรือ False Negative คือ จำนวนข้อมูลที่มีการจำแนกคลาสเป้าหมาย Positive เป็น Negative

ส่วนโมเดลการยืนยันตัวตนผู้ใช้งานทำการประเมินประสิทธิภาพโดยใช้ค่าความถูกต้อง (Accuracy) ค่าการปฏิเสธผิด (False Rejection Rate: FRR) หมายถึง ค่าอัตราการปฏิเสธการผ่านแก่ผู้ใช้ที่ถูกต้อง และค่าการอนุญาตผิด (False Acceptance Rate: FAR) หมายถึง ค่าอัตราการหลุดรอดของผู้แอบอ้างจากการตรวจจับ ซึ่งสามารถคำนวณได้จากสมการ

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme [3]

ผู้แต่ง : ang Hoang, Deokjai Choi, และ Thuc Nguyen

Thang Hoang, Deokjai Choi, และ Thuc Nguyen ได้กล่าวถึงระบบการยืนยันตัวตนโดยใช้การเดิน จาก inertial sensors ที่อยู่บนสมาร์ทโฟน โดยเป็นระบบการยืนยันตัวตนด้วยการเดินบนสมาร์ทโฟนที่มี ความปลอดภัยและความเป็นส่วนตัว โดยใช้ประโยชน์จาก fuzzy commitment scheme แทนที่จะเก็บ แม่แบบ(template)การเดินดั้งเดิม สำหรับการยืนยันผู้ใช้เช่นเดียวกับวิธีทั่วไป ผู้วิจัยได้ตรวจสอบผู้ใช้ผ่านคีย์ ที่เก็บไว้ซึ่งมีการเข้ารหัสไบโอเมตริก โดยแม่แบบการเดินที่รวบรวมจากเซนเซอร์ accelerometer ของ สมาร์ทโฟน ซึ่งการจำแนกแม่แบบ (template) การเดินจะถูกตรวจสอบเพื่อกำหนดค่าพารามิเตอร์ที่เหมาะสม เพื่อสร้างระบบเข้ารหัสไบโอเมตริกที่อิงกับการเดิน ซึ่งประสิทธิภาพของระบบที่ผู้วิจัยนำเสนอนั้นได้ทำการประเมินบนชุดข้อมูลรวมถึงสัญญาณการเดินของอาสาสมัคร 34 คน ผลลัพธ์ที่ได้คือ ได้ค่า False Acceptance Rate(อัตราการปฏิเสธการผ่านแก่ผู้ใช้ที่ถูกต้อง) 0% และค่า False Rejection Rate (อัตราการหลุดรอดของผู้แปลกปลอมจากการตรวจจับ) 16.18% สอดคล้องกับความยาวคีย์ของระบบ ความปลอดภัยที่ระดับ 139 บิต ผลจากการทดลองแสดงให้เห็นว่าค่าของ False Rejection Rate ยังคงค่อนข้างสูงเมื่อเทียบกับการเข้ารหัสไบโอเมตริกอื่น(การใช้ลายนิ้วมือ ม่านตา และลายเซ็น)

2.2.2 Multi-sensor authentication to improve smartphone security [4]

ผู้แต่ง : Wei-Han Lee และ Ruby B. Lee

Wei-Han Lee และ Ruby B. Lee ได้ใช้เซนเซอร์ 3 ตัวคือ accelerometer, orientation sensor และ magnetometer โดยใช้เทคนิค SVM เป็นอัลกอริทึมการจำแนกในระบบ ซึ่งข้อมูลที่ใช้นั้นมาจากการเก็บด้วย ตัวเอง(นักเรียนที่จบการศึกษาแล้ว 4 คน จากมหาวิทยาลัย Princeton) และข้อมูลจาก Kayacık et al., 2014 (เจ้าหน้าที่และนักเรียน 4 คน จากมหาวิทยาลัย Glasgow Caledonian) เพื่อแยกแยะเจ้าของสมาร์ทโฟนกับ ผู้ใช้รายอื่นซึ่งอาจเป็นผู้โจมตีหรือขโมย แล้วนำผลลัพธ์ของค่า accuracy จากการใช้อัลกอริทึม SVM มาทำการ เปรียบเทียบกันของแต่ละเซนเซอร์ กับการรวมกันของ 2 เซนเซอร์ และ 3 เซนเซอร์ ซึ่งพบว่า orientation sensor นั้นมีค่า accuracy น้อยกว่า accelerometer และ magnetometer ดังนั้น ข้อมูลที่ได้มาจาก orientation sensor จึงไม่สำคัญเท่ากับ accelerometer และ magnetometer ซึ่งมีแนวโน้มที่จะวัดได้เสถียรมากกว่า

2.2.3 Smartphone-Based Gait Recognition From Authentication to Imitation [5]

ผู้แต่ง : Muhammad Muaaz และ René Mayrhofer

Muhammad Muaaz และ René Mayrhofer กล่าวถึงการประเมินความปลอดภัย ของระบบการจดจำการเดิน โดยใช้สมาร์ทโฟนและได้มีการพัฒนา แอปพลิเคชัน Android ซึ่งใช้เซนเซอร์ accelerometer ของสมาร์ทโฟนเพื่อเก็บข้อมูลการเดินอย่างต่อเนื่อง และได้ทดสอบประสิทธิภาพของระบบนี้ โดยใช้ชุดข้อมูลจำนวน 35 คน แล้วจึงให้ผู้อื่นลองปลอมตัวเป็นผู้ใช้สมาร์ทโฟนแทน และได้เปรียบเทียบกับ template การเดินที่บันทึกไว้แล้วโดยใช้ DTW(Dynamic Time Warping) เป็นการวัดระยะทาง ผลลัพธ์คือ มีความผิดพลาด 13% ซึ่งหมายความว่าระบบยืนยันตัวตนนั้นมีความถูกต้องถึง 87%

บทที่ 3
วิธีการดำเนินงานศึกษาค้นคว้า

การพัฒนาโมเดลการรู้จำอิริยาบถและการยืนยันตัวตนผู้ใช้งานโดยใช้การการเรียนรู้ของเครื่องจักร (Machine learning) ในโครงงานนี้ผู้จัดทำได้ศึกษาและดำเนินงานตามขั้นตอน ดังต่อไปนี้

3.1 วิธีการดำเนินการวิจัย

1. Download dataset จาก https://www.utwente.nl/en/eemcs/ps/research/dataset
    a. Physical activity recognition dataset
    b. Sensors activity dataset
2. สร้างโมเดลสำหรับการรู้จำอิริยาบถโดยใช้ Physical activity recognition dataset และ classification algorithms.
3. ทดสอบประสิทธิภาพของโมเดลการรู้จำอิริยาบถ
4. สร้างโมเดลสำหรับการยืนยันตัวตนผู้ใช้งานโดยใช้ Sensors activity dataset และ One-class classifier algorithms
5. ทดสอบประสิทธิภาพโมเดลสำหรับการยืนยันตัวตนผู้ใช้งาน
6. พัฒนาแอปพลิเคชันแอนดรอยด์สำหรับการยืนยันตัวตน
7. สรุปผลการทดลองและจัดทำรูปเล่มโครงงาน

3.2 แผนการดำเนินงานตลอดโครงงาน

ตารางที่ 1 แผนการดำเนินงาน

ลำดับที่	แผนการดำเนินงาน	2561					2		
		ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี
1	ศึกษางานวิจัยและทฤษฎีเกี่ยวกับการรู้จำอิริยาบถ และ การยืนยันตัวตนโดยใช้การรู้จำอิริยาบถ	x	x						

2	ศึกษาหาชุดข้อมูลและเลือกอัลกอริทึมเพื่อนำมาใช้ในการวิจัย ดาวน์โหลดชุดข้อมูล จาก https://www.utwente.nl/en/eemcs/ps/research/dataset/ ● Physical activity recognition dataset ● Sensors activity dataset			x	x				
3	ทำการสกัดคุณลักษณะ (Feature extraction) โดยใช้ Physical activity recognition dataset และทำการสร้างโมเดลสำหรับการรู้จำอิริยาบถ					x	x		
4	ทดสอบประสิทธิภาพของโมเดลการรู้จำอิริยาบถ						x		

ลำดับที่	แผนการดำเนินงาน	2561					2562				
		ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
5	ทำการสกัดคุณลักษณะ (Feature extraction) โดยใช้ Sensors activity dataset ของแต่ละผู้ใช้งานในชุดข้อมูลเพื่อนำไปใช้ในการเปรียบเทียบกับคุณลักษณะที่ได้จากเซนเซอร์ของสมาร์ทโฟน (Feature matching)							x			

6	ยืนยันตัวตนผู้ใช้งานโดยใช้เกณฑ์การให้คะแนน (Scoring model)							x	x		
7	ทดสอบประสิทธิภาพการยืนยันตัวตนผู้ใช้งานในชุดข้อมูล								x		
8	พัฒนาแอปพลิเคชันต้นแบบแอนดรอยด์สำหรับการยืนยันตัวตน								x	x	
9	สรุปผลการทดลองและจัดทำรูปเล่มโครงงาน									x	

3.3 อุปกรณ์และเครื่องมือที่ใช้ในการวิจัย

3.3.1 ฮาร์ดแวร์

1. ระบบปฏิบัติการ: Windows 10 Education (64-bit)
2. หน่วยประมวลผล: Intel Core i5-3470 @ 3.20 GHz
3. หน่วยความจำ: RAM 8 GB

3.3.2 ซอฟต์แวร์

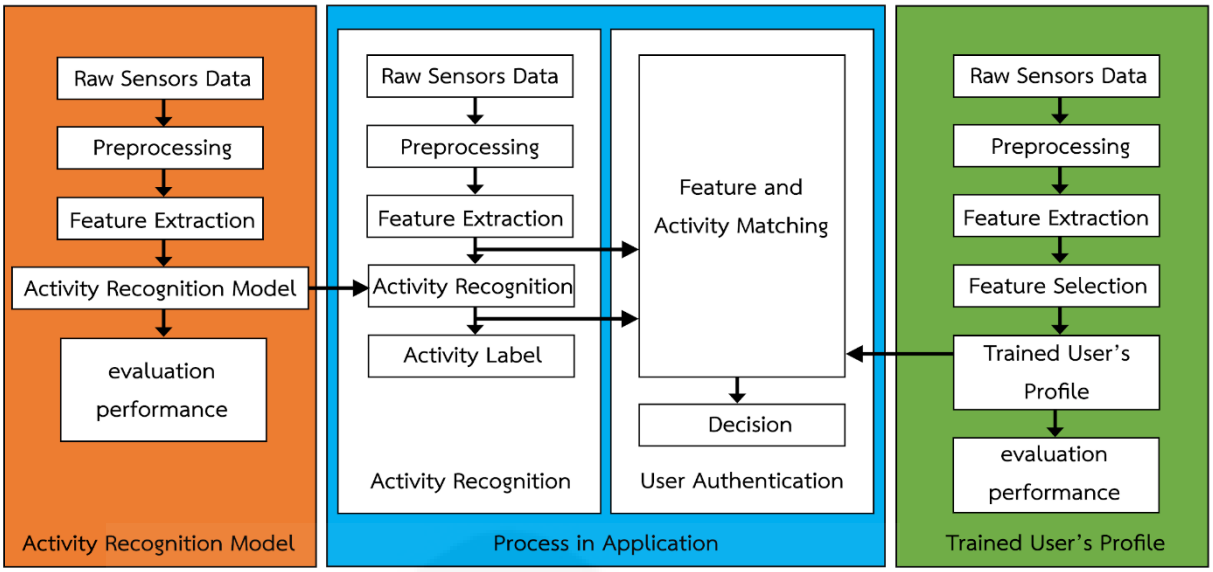
1. Anaconda (Jupyter Notebook)
2. Android studio

3.3.2 ภาษาที่ใช้

1. Python สำหรับการเตรียมข้อมูล (Data pree-processing) และพัฒนาโมเดล (Modeling)
2. Java สำหรับพัฒนาแอปพลิเคชันสำหรับการยืนยันตัวตนผู้ใช้งาน

3.4 การออกแบบและพัฒนา

วิธีการพัฒนาแบ่งออกเป็น 3 ส่วน ประกอบด้วย 1. การสร้างโมเดลสำหรับการรู้จำอิริยาบถ 2. การสร้างโมเดลสำหรับการยืนยันตัวตนผู้ใช้งาน และ 3. การนำโมเดลที่ได้ไปใช้ในการพัฒนาแอปพลิเคชัน รายละเอียด ดังภาพที่ 9



ภาพที่ 9 แผนภาพแสดงขั้นตอนการทำงานของระบบ

จากภาพที่ 9 ในส่วนของด้านซ้าย คือขั้นตอนการสร้างโมเดลสำหรับการรู้จำอิริยาบถ ประกอบด้วย การนำข้อมูลดิบมาทำการเตรียมข้อมูล ได้แก่ การลดสัญญาณรบกวน และการสกัดคุณลักษณะ จากนั้นทำ การสร้างโมเดล และทำการประเมินโมเดล ส่วนของด้านขวา คือ ขั้นตอนการสร้างโมเดล สำหรับการ ยืนยันตัวตนผู้ใช้งานได้ดำเนินงานคล้ายขั้นตอนการสร้างโมเดลสำหรับการรู้จำอิริยาบถ เพียงแต่เพิ่มขั้นตอน การเลือกคุณลักษณะที่ดำเนินการต่อจากการสกัดคุณลักษณะ และส่วนกลางของภาพ คือ การนำโมเดล ที่ได้มาใช้สำหรับพัฒนาแอปพลิเคชันสำหรับการยืนยันตัวตนผู้ใช้งาน โดยหลักการทำงานภายในแอปพลิเคชัน จะเหมือนกับขั้นตอนของการสร้างโมเดล

3.4.1 ชุดข้อมูล (Dataset)

1. Physical activity recognition dataset

ชุดข้อมูลนี้เก็บข้อมูลจากผู้เข้าร่วมทั้งหมด 4 คนเป็นเพศชายอายุ 25 - 30 ปี ทำอิริยาบถทั้งหมดจำนวน 6 อิริยาบถ คือ การเดิน การนั่ง การยืน การวิ่ง การเดินขึ้นบันได และการเดินลงบันได โดยให้ผู้เข้าร่วมทำแต่ละอิริยาบถประมาณ 3-4 นาที ตำแหน่ง ของสมาร์ทโฟน ได้แก่ กระเป๋ากางเกงขวา ต้นแขนขวา ข้อมือขวา และเอว เก็บข้อมูลจากเซนเซอร์ทั้งหมด 3 เซนเซอร์ได้แก่ เซนเซอร์วัดค่าความเร่ง เซนเซอร์วัด การหมุน และเซนเซอร์วัดสนามแม่เหล็กอัตราการเก็บข้อมูลอยู่ที่ 50 Hz. หรือ 50 ตัวอย่างต่อวินาที ภายในชุดข้อมูลประกอบไปด้วยไฟล์ csv ทั้งหมด 4 ไฟล์ได้แก่ Arm.csv, Belt.csv, Wrist.csv, Pocket.csv

โดยภาพที่ 10 แสดงข้อมูลภายในไฟล์ที่เก็บรวบรวมข้อมูลจากทั้ง 3 เซนเซอร์ ได้แก่

1. Timestamp คือ เวลาการเก็บข้อมูลตัวอย่าง

2. Ax คือ ค่าจากแกน x ของเซนเซอร์วัดความเร่ง (Accelerometer)

3. Ay คือ ค่าจากแกน y ของเซนเซอร์วัดความเร่ง (Accelerometer)

4. Az คือ ค่าจากแกน z ของเซนเซอร์วัดความเร่ง (Accelerometer)

5. Gx คือ ค่าจากแกน x ของเซนเซอร์วัดการหมุน (Gyroscope)

6. Gy คือ ค่าจากแกน y ของเซนเซอร์วัดการหมุน (Gyroscope)

7. Gz คือ ค่าจากแกน z ของเซนเซอร์วัดการหมุน (Gyroscope)

8. Mx คือ ค่าจากแกน x ของเซนเซอร์วัดสนามแม่เหล็ก (Magnetometer)

9. My คือ ค่าจากแกน y ของเซนเซอร์วัดสนามแม่เหล็ก (Magnetometer)

10. Mz คือ ค่าจากแกน z ของเซนเซอร์วัดสนามแม่เหล็ก (Magnetometer)

11. Activity Label คือ ประเภทอิริยาบถ

Time_Stamp	Ax	Ay	Az	Gx	Gy	Gz	Mx	My	Mz	Activity_Label
1364400000000	0.23154591	-9.32994	-3.02372	0.275195	-0.11179	0.019242	-19.68	29.64	-6.6	Downstairs
1364400000020	1.0487667	-8.85323	-1.22583	0.80115	1.128879	0.011912	-19.619999	29.4	-6.72	Downstairs
1364400000040	0.939804	-9.507	-2.04305	0.807258	1.709201	-0.03238	-19.68	29.34	-6.96	Downstairs
1364400000060	0.6401563	-9.83389	-2.15202	0.602008	2.062892	0.003665	-19.859999	29.22	-7.14	Downstairs
1364400000080	0.027240695	-9.24822	-1.32117	0.629802	2.191174	0.159741	-19.68	29.039999	-7.98	Downstairs
1364400000100	0.47671217	-9.34356	-1.45738	0.718988	1.960267	0.295659	-19.5	28.92	-8.4	Downstairs
1364400000120	1.0896279	-9.61597	-0.8036	0.718683	1.903456	0.347277	-19.32	28.74	-8.76	Downstairs
1364400000140	1.6889231	-9.45252	0.40861	0.725403	2.049147	0.331394	-18.779999	28.74	-9.78	Downstairs
1364400000160	2.4380422	-9.60235	0.612916	0.627359	2.261729	0.308792	-18.48	28.74	-10.44	Downstairs

ภาพที่ 10 ตัวอย่างข้อมูลจากไฟล์ Pocket.csv

2. Sensors activity dataset

ชุดข้อมูลนี้เก็บข้อมูลจากผู้เข้าร่วมทั้งหมด 10 คนเป็นเพศชายอายุ 25 - 30 ปี ทำอิริยาบถทั้งหมดจำนวน 6 อิริยาบถ คือ การเดิน การนั่ง การยืน การวิ่ง การเดินขึ้นบันได และการเดินลงบันได โดยให้ผู้เข้าร่วมทำแต่ละอิริยาบถประมาณ 3-4 นาที ตำแหน่งของ สมาร์ทโฟน ได้แก่ กระเป๋ากางเกงขวา กระเป๋ากางเกงซ้าย ต้นแขนขวา ข้อมือขวา และเอว เก็บข้อมูลจากเซนเซอร์ทั้งหมด 4 เซนเซอร์ได้แก่ เซนเซอร์วัดค่าความเร่ง เซนเซอร์วัดค่า ความเร่งเชิงเส้น เซนเซอร์วัดการหมุน และเซนเซอร์วัดสนามแม่เหล็ก อัตราการเก็บข้อมูล อยู่ที่ 50 Hz. หรือ 50 ตัวอย่างต่อวินาที ภายในชุดข้อมูลประกอบไปด้วยไฟล์ csv ทั้งหมด 10 ไฟล์ได้แก่ Participant_1.csv จนถึง Participant_10.csv

โดยภาพที่ 11 แสดงข้อมูลภายในไฟล์ที่เก็บรวบรวมข้อมูลจากทั้ง 4 เซนเซอร์ได้แก่

1. Timestamp คือ เวลาการเก็บข้อมูลตัวอย่าง
2. Ax คือ ค่าจากแกน x ของเซนเซอร์วัดความเร่ง (Accelerometer)
3. Ay คือ ค่าจากแกน y ของเซนเซอร์วัดความเร่ง (Accelerometer)
4. Az คือ ค่าจากแกน z ของเซนเซอร์วัดความเร่ง (Accelerometer)
2. Lx คือ ค่าจากแกน x ของเซนเซอร์วัดความเร่งเชิงเส้น (Linear accelerator)
3. Ly คือ ค่าจากแกน y ของเซนเซอร์วัดความเร่งเชิงเส้น (Linear accelerator)
4. Lz คือ ค่าจากแกน z ของเซนเซอร์วัดความเร่งเชิงเส้น (Linear accelerator)
5. Gx คือ ค่าจากแกน x ของเซนเซอร์วัดการหมุน (Gyroscope)
6. Gy คือ ค่าจากแกน y ของเซนเซอร์วัดการหมุน (Gyroscope)
7. Gz คือ ค่าจากแกน z ของเซนเซอร์วัดการหมุน (Gyroscope)
8. Mx คือ ค่าจากแกน x ของเซนเซอร์วัดสนามแม่เหล็ก (Magnetometer)
9. My คือ ค่าจากแกน y ของเซนเซอร์วัดสนามแม่เหล็ก (Magnetometer)
10. Mz คือ ค่าจากแกน z ของเซนเซอร์วัดสนามแม่เหล็ก (Magnetometer)
11. Activity Label คือ ประเภทอิริยาบถ
12. ระบุตำแหน่งของสมาร์ทโฟนไว้ที่มุมซ้ายบนของไฟล์

Belt													
time_stamp	Ax	Ay	Az	Lx	Ly	Lz	Gx	Gy	Gz	Mx	My	Mz	
1390000000000	4.3177	-2.3699	-0.43585	-5.3574	-0.95749	0.31886	0.52565	-0.18815	0.28588	-22.2	6.48	4.56	walking
1390000000000	5.1213	-2.1929	-0.70826	-4.5352	-0.67629	0.081656	0.43616	0.16371	0.19823	-21.9	7.14	5.34	walking
1390000000000	5.6116	-2.1248	-1.2667	-4.0408	-0.55974	-0.52369	0.32284	0.43127	0.21197	-21.6	7.5	5.64	walking
1390000000000	6.0611	-1.471	-2.4789	-3.5908	0.15795	-1.8802	0.20403	0.72846	0.16493	-21.12	8.4	6.12	walking
1390000000000	7.7091	-0.77636	-3.5413	-1.945	0.85796	-2.9953	0.09896	0.94379	0.064752	-20.94	8.7	6.3	walking
1390000000000	11.21	-0.98067	-1.6753	1.543	0.63528	-1.3342	0.18326	0.81215	-0.03329	-20.7	9.3	6.42	walking
1390000000000	12.612	-1.5119	0.98067	2.94	0.083211	1.2442	0.58857	0.52046	-0.25137	-20.64	9.66	6.3	walking
1390000000000	15.854	-5.8567	2.2337	5.1271	-2.5343	3.3763	0.6466	0.35125	-0.25992	-20.7	10.02	6.18	walking
1390000000000	17.243	-4.6037	-10.12	7.5748	-2.971	-9.9641	0.048869	0.31551	0.35613	-20.76	10.62	5.82	walking

ภาพที่ 11 ตัวอย่างข้อมูลจากไฟล์ Participant_1.csv

3.4.2 การรู้จำอิริยาบถ (Activity Recognition)

จากภาพที่ 9 ส่วนด้านซ้ายการทำเหมืองข้อมูลการรู้จำอิริยาบถ ถือเป็นปัญหา การจำแนกประเภทด้วยกลไกการเรียนรู้ด้วยข้อมูลสำคัญจากคุณลักษณะที่สกัดจากสัญญาณ จากนั้นจึงทำการจำแนกตามคุณลักษณะอิริยาบถซึ่งอัลกอริทึมที่จะนำมาจำแนกประเภทอิริยาบถประกอบด้วย 4 อัลกอริทึมคือ Decision Tree, K-Nearest Neighbors, Support Vector

Machine และ Bayesian networks เพื่อให้สามารถจำแนกอิริยาบถที่แตกต่างกัน ทั้งหมด 6 อิริยาบถ ได้แก่ การเดิน การนั่ง การยืน การวิ่ง การเดินขึ้นบันได และการเดินลงบันได

ชุดข้อมูลที่ใช้คือ Physical activity recognition dataset

1. การเตรียมข้อมูล (Data Pre-Processing)

1.1 การกำจัดข้อมูลที่ไม่สมบูรณ์ออกจากข้อมูลดิบโดยใช้วิธี Average smoothing

ในโครงงานนี้ใช้จำนวนข้อมูลสองจุดที่ติดกัน เพื่อให้ข้อมูลไม่แตกต่างกันเกินไป ซึ่งได้ใช้คำสั่ง ด้านล่างนี้ในการดำเนินการ

```
smooth = df_1.rolling(3).mean()
```

	Time_Stamp	Ax	Ay	Az	Gx	Gy	Gz	Mx	My	Mz	Activity_Label
0	1364400000000	0.740039	-9.230056	-2.097534	0.627868	0.908764	-0.000407	-19.660000	29.460000	-6.760000	Downstairs
1	1364400000020	0.876242	-9.398040	-1.806966	0.736805	1.633657	-0.005600	-19.719999	29.320000	-6.940000	Downstairs
2	1364400000040	0.535734	-9.529703	-1.838747	0.679689	1.987756	0.043677	-19.740000	29.200000	-7.360000	Downstairs
3	1364400000060	0.381370	-9.475222	-1.643522	0.650266	2.071444	0.153022	-19.680000	29.060000	-7.840000	Downstairs
4	1364400000080	0.531194	-9.402580	-1.194050	0.689158	2.018299	0.267559	-19.500000	28.900000	-8.380000	Downstairs
5	1364400000100	1.085088	-9.470682	-0.617456	0.721025	1.970957	0.324777	-19.200000	28.800000	-8.980000	Downstairs
6	1364400000120	1.738864	-9.556944	0.072642	0.690481	2.071444	0.329155	-18.860000	28.740000	-9.660000	Downstairs

ภาพที่ 12 ตัวอย่างข้อมูลจากไฟล์ Pocket.csv หลังการกำจัดข้อมูลที่ไม่สมบูรณ์ด้วยวิธี Average smoothing

หลังจากการกำจัดข้อมูลที่ไม่สมบูรณ์ด้วยวิธี Average smoothing แล้ว จะได้ข้อมูลของแต่ละเซนเซอร์ที่มีความใกล้เคียงกัน และมีความสมบูรณ์กว่าเดิมดังภาพที่ 11

1.2 การคำนวณและเพิ่มค่า magnitude ของแต่ละเซนเซอร์ เนื่องจากความไวของเซนเซอร์มีผลต่อการเปลี่ยนทิศทางของสมาร์ทโฟนจึงทำการคำนวณและเพิ่มค่า magnitude ของแต่ละเซนเซอร์เพื่อแก้ไขปัญหานี้ โดยใช้สมการ

โดยฎีค่า x, y และ z คือความที่ได้จากแกน x, y และ z ตามลำดับจากเซนเซอร์

```
df_smoothing['Mag_A'] = np.sqrt(np.power(df_smoothing['Ax'],2)+np.power(df_smoothing['Ay'],2)+np.power(df_smoothing['Az'],2))
df_smoothing['Mag_G'] = np.sqrt(np.power(df_smoothing['Gx'],2)+np.power(df_smoothing['Gy'],2)+np.power(df_smoothing['Gz'],2))
df_smoothing['Mag_M'] = np.sqrt(np.power(df_smoothing['Mx'],2)+np.power(df_smoothing['My'],2)+np.power(df_smoothing['Mz'],2))
```

	Time_Stamp	Ax	Ay	Az	Gx	Gy	Gz	Mx	My	Mz	Mag_A	Mag_G	Mag_M	Activity_Label
0	1364400000000	0.740039	-9.230056	-2.097534	0.627868	0.908764	-0.000407	-19.660000	29.46	-6.76	9.494274	1.104568	36.056966	Downstairs
1	1364400000020	0.876242	-9.398040	-1.806966	0.736805	1.633657	-0.005600	-19.719999	29.32	-6.94	9.610207	1.792136	36.009782	Downstairs
2	1364400000040	0.535734	-9.529703	-1.838747	0.679689	1.987756	0.043677	-19.740000	29.20	-7.36	9.720249	2.101204	36.006627	Downstairs
3	1364400000060	0.381370	-9.475222	-1.643522	0.650266	2.071444	0.153022	-19.680000	29.06	-7.84	9.624263	2.176498	35.961807	Downstairs
4	1364400000080	0.531194	-9.402580	-1.194050	0.689158	2.018299	0.267559	-19.500000	28.90	-8.38	9.492968	2.149432	35.856441	Downstairs

ภาพที่ 13 ตัวอย่างข้อมูลจากไฟล์ Pocket.csv หลังเพิ่มค่า magnitude

หลังจากทำการคำนวณและเพิ่มค่า magnitude แล้ว จะได้ค่าของคุณลักษณะเพิ่มขึ้น 3 ค่า คือ ค่าขนาดของแต่ละเซนเซอร์ดังภาพที่ 12

1.3 การสกัดคุณลักษณะ (Feature extraction) คุณลักษณะที่เลือกได้แก่ max, min, variance, mean, skewness และ kurtosis แบ่งข้อมูลเท่ากับ 5 วินาที จากนั้น ทำการหาคุณลักษณะของแต่ละกลุ่มข้อมูล (Data segment) เพื่อนำคุณลักษณะที่ได้นั้น มาทำการสร้างโมเดลการรู้จำอิริยาบถ ซึ่งมีข้อมูลดังภาพที่ 13

User_id	Ax_max	Ay_max	Az_max	Gx_max	Gy_max	Gz_max	Mag_A_max	Mag_G_max	Mag_M_max	...	Az_kurt	Gx_kurt	Gy_kurt	Gz_kurt
1	6.515067	-1.298463	5.9566	4.818933	6.883000	3.074600	21.054306	6.932296	72.099515	...	0.852710	-0.677570	-0.270184	-0.323235
1	8.503500	-0.513017	7.8408	4.647867	6.981667	3.123067	22.928676	7.061921	72.611553	...	0.185435	-0.312378	0.428971	-0.465484
1	6.415400	3.781907	3.8591	5.346800	5.667133	3.442800	20.216861	5.977822	112.352136	...	-0.556158	-0.087942	0.097265	0.022201
1	6.678500	0.077167	3.8001	4.785233	6.021100	2.966067	22.275976	6.375510	62.734195	...	-0.569975	-0.599718	0.081225	-0.290344
1	8.349333	4.812533	4.5628	5.236933	5.729100	2.997633	20.012457	5.913936	83.716921	...	-0.581514	-0.467823	-0.113350	-0.466400

ภาพที่ 14 ตัวอย่างข้อมูลหลังจากทำการสกัดคุณลักษณะ

หลังจากทำการเตรียมข้อมูล (Data Pre-Processing) ตามขั้นตอนข้างต้นแล้ว จะได้ข้อมูลคุณลักษณะของแต่ละเซนเซอร์ เพื่อที่จะนำไปทำการสร้างโมเดลการรู้จำอิริยาบถ

2. สร้างโมเดลการรู้จำอิริยาบถ (Activity recognition model) ด้วย Python

การรู้จำอิริยาบถเป็นปัญหาการจำแนกประเภทของคลาสหลายคลาส โดยจากการ สำรวจอัลกอริทึมการเรียนรู้ของเครื่องที่ใช้ในการจำแนกที่มีความนิยมในการใช้งานที่สุด ได้แก่ Support Vector Machines k-Nearest Neighbors Decision Tree และ Bayesian networks (Gaussian Naive Bayes) ซึ่งอัลกอริทึมประเภทนี้มีความเหมาะสม ที่จะนำไปใช้ ในสมาร์ทโฟนเพราะใช้หน่วยความจำในการนคำนวณน้อย [6]

2.1 data frame of features

สร้าง data frame สำหรับคุณลักษณะที่ได้จากการสกัดคุณลักษณะ โดยไม่ได้ใช้ ข้อมูลในส่วนของ Timestamp (column index: 0) ดังคำสั่งด้านล่าง

```
#feature dataframe
df_feat = pd.DataFrame(Pocket, columns = Pocket.columns[1:-1])
```

2.2 define feature and label

กำหนดค่าของตัวแปร x และ y เท่ากับ ชุดข้อมูลของคุณลักษณะ และคลาสตามลำดับ เพื่อใช้สำหรับการสร้างโมเดลการรู้จำอิริยาบถโดยใช้คำสั่งด้านล่าง

```
X = df_feat
y = Pocket.Activity_Label
```

2.3 Classifying algorithms

สำหรับคำสั่งด้านล่างนั้นเป็นกาใช้ Classifying algorithm ในสร้างโมเดลการรู้จำอิริยาบถ

```
knn = KNeighborsClassifier(n_neighbors=1)
svc = svm.SVC(kernel='linear')
tree = DecisionTreeClassifier()
gnb = GaussianNB()
```

3.4.3 การยืนยันตัวตนผู้ใช้งานจากการรู้จำอิริยาบถ (User authentication for activity recognition)

ในขั้นตอนการยืนยันตัวตนระบบจะตัดสินใจว่าผู้ใช้นั้นเป็นผู้ใช้ที่ถูกต้องหรือเป็นผู้แอบอ้าง ในสถานการณ์จริงตัวอย่างข้อมูลที่ได้รับจากเซนเซอร์สมาร์ทโฟนมาจากผู้ใช้ที่ถูกต้องโดยส่วนใหญ่ ในขณะที่มีตัวอย่างไม่กี่คนจากผู้แอบอ้าง ดังนั้นเราจึงถือว่าการยืนยันตัวตนผู้ใช้งานเป็นปัญหาการจำแนก แบบคลาสเดียว (One-class classification) คือเป็นหรือไม่เป็นเจ้าของสมาร์ทโฟน โดยวิธีการที่เลือกใช้ คือ

- One-class SVM
- One-Class K-Nearest Neighbor

1. การสกัดคุณลักษณะ (Feature extraction)

คุณลักษณะที่เลือกได้แก่ max, min, variance, mean, skewness, kurtosis, peak แบ่งข้อมูลเท่ากับ 5 วินาที จากนั้นทำการหาคุณลักษณะของแต่กลุ่มข้อมูล (Data segment)

User_id	Ax_max	Ay_max	Az_max	Gx_max	Gy_max	Gz_max	Mag_A_max	Mag_G_max	Mag_M_max	...	Az_kurt	Gx_kurt	Gy_kurt	Gz_kurt
1	6.515067	-1.298463	5.9566	4.818933	6.883000	3.074600	21.054306	6.932296	72.099515	...	0.852710	-0.677570	-0.270184	-0.323235
1	8.503500	-0.513017	7.8408	4.647867	6.981667	3.123067	22.928676	7.061921	72.611553	...	0.185435	-0.312378	0.428971	-0.465484
1	6.415400	3.781907	3.8591	5.346800	5.667133	3.442800	20.216861	5.977822	112.352136	...	-0.556158	-0.087942	0.097265	0.022201
1	6.678500	0.077167	3.8001	4.785233	6.021100	2.966067	22.275976	6.375510	62.734195	...	-0.569975	-0.599718	0.081225	-0.290344
1	8.349333	4.812533	4.5628	5.236933	5.729100	2.997633	20.012457	5.913936	83.716921	...	-0.581514	-0.467823	-0.113350	-0.466400

ภาพที่ 15 การสกัดคุณลักษณะ

2. Feature Selection

คำสั่งสำหรับใช้ในการทำ Feature Selection ด้วย ReliefF คือ

```
features, labels = data.drop('User_id', axis=1).values,
                   data['User_id'].values
X_train, X_test, y_train, y_test =
                    train_test_split(features, labels)
fs = ReliefF(n_features_to_select=20)
fs.fit(X_train, y_train)
```

3. การยืนยันตัวตน

3.1 การยืนยันตัวตนด้วย one-class svm

```
user = df_New_features.loc[df_New_features.User_id==1]
impostor = df_New_features.loc[df_New_features.User_id!=1]

train_feature = user.loc[0:26, :]
train_feature = user.drop('User_id', 1)

X_test_1 = user.iloc[26:, :].drop('User_id',1)
X_test_2 = impostor.drop('User_id',1)
X_test = X_test_1.append(X_test_2)

oneclass = svm.OneClassSVM(kernel='rbf', gamma=0.001, nu=0.025)

Y_1 = user.loc[26:, 'User_id']
Y_2 = impostor['User_id']
Y_test= Y_1.append(Y_2)
```

```
oneclass.fit(train_feature)
fraud_pred = oneclass.predict(X_test)
```

ทำการแบ่งข้อมูลจากชุดข้อมูล (Dataset) ออกเป็น 2 กลุ่ม ประกอบด้วย กลุ่ม ของผู้ใช้ที่ถูกต้อง 1 คน และกลุ่มผู้ใช้แอบอ้าง 9 คน และทำการสร้างโมเดลสำหรับ การยืนยันตัวตนโดยใช้ข้อมูล 80% จากกลุ่มผู้ใช้ที่ถูกต้อง จากนั้นทำการทดสอบโมเดล โดยใช้ข้อมูลที่เหลืออีก 20% จากกลุ่มผู้ใช้ที่ถูกต้องและข้อมูลทั้งหมดของกลุ่มผู้ใช้แอบอ้าง

3.2 การยืนยันตัวตนด้วย one-class k-NN

คำสั่งสำหรับการยืนยันตัวตนด้วย one-class k-NN

```
# find th each of rows train data user1
th_train_user1 = []
for i in range(len(train_data_user1)):
    dict1 = {}
    check_min = []
    if(i == 0):
        for j in range(len(train_data_user1)):
            dict2 = {}
            x = train_data_user1.iloc[i]
            y = train_data_user1.iloc[j]
            distance = math.sqrt(sum([(a - b) ** 2 for a, b in zip(x, y)]))
            if(distance != 0.0):
                dict2.update({'distance': distance})
                check_min.append(dict2)

    else:
        for j in range(len(train_data_user1)):
            dict2 = {}
            x = train_data_user1.iloc[i]
            y = train_data_user1.iloc[j]
            distance = math.sqrt(sum([(a - b) ** 2 for a, b in zip(x, y)]))
            if(distance != 0.0):
                dict2.update({'distance': distance})
                check_min.append(dict2)

    s = pd.DataFrame(check_min)
    cmin = s.sort_values(by=['distance'])
    min_k10 = cmin[1:11].mean()
    dict1.update(cmin.min()/min_k10)
    th_train_user1.append(dict1)
```

ทำการหาข้อมูลตัวที่ใกล้ที่สุดของข้อมูลแต่ละตัวของข้อมูลฝึกสอน จากนั้นจะนำ ค่าที่ใกล้ที่สุด 10 ค่า มาหาค่า Treshold เพื่อที่จะนำไปเปรียบเทียบกับค่า Treshold ของข้อมูลทดสอบ

```
# find th each of rows test data user1
th_test_user1 = []
for i in range(len(test_data_user1)):
    dict1 = {}
    check_min = []

    for j in range(len(train_data_user1)):
        dict2 = {}
        x = test_data_user1.iloc[i]
        y = train_data_user1.iloc[j]
        distance = math.sqrt(sum([(a - b) ** 2 for a, b in zip(x, y)]))
        dict2.update({'distance': distance})
        check_min.append(dict2)

    s = pd.DataFrame(check_min)
    cmin = s.sort_values(by=['distance'])
    min_k10 = cmin[1:11].mean()
    dict1.update(cmin.min()/min_k10)
    th_test_user1.append(dict1)
```

ทำการหาข้อมูลตัวที่ใกล้ที่สุดของข้อมูลแต่ละตัวของข้อมูลทดสอบ จากนั้นจะนำค่าที่ใกล้ที่สุด 10 ค่า มาหาค่า Threshold

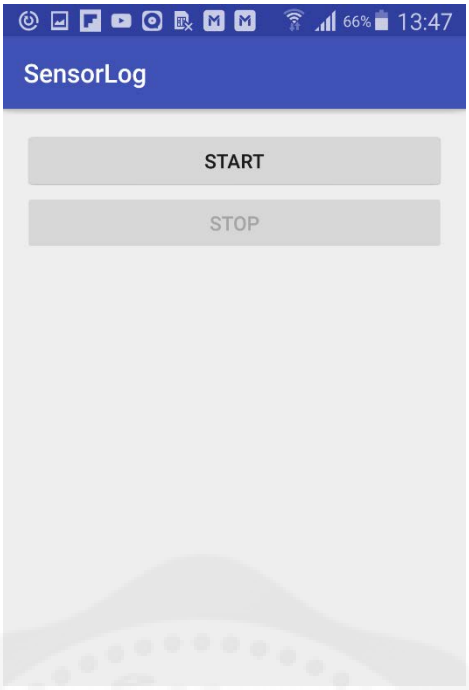
```
# test data user1
r_test_user1 = []
yes = 0
no = 0
for i in range(len(test_data_user1)):
    dict1 = {}
    r = pd.DataFrame(th_test_user1).iloc[i]
    if(r.item() < th_user1.item()):
        y = 'Yes'
        dict1.update({'Result': y})
        r_test_user1.append(dict1)
        yes += 1
    else:
        n = 'No'
        dict1.update({'Result': n})
        r_test_user1.append(dict1)
        no += 1
```

นำค่า Treshold ของแต่ละตัวมาเปรียบเทียบกับค่า Treshold ของข้อมูลฝึกสอน จะได้ผลลัพธ์ออกมาว่าเป็นหรือไม่เป็นผู้ใช้งาน

3.4.4 การเก็บข้อมูลเพื่อใช้สำหรับยืนยันตัวตนผู้ใช้งาน

เก็บข้อมูลจากการทำอิริยาบถทั้งหมดจำนวน 6 อิริยาบถ คือ การเดิน การนั่ง การยืน การวิ่ง การเดินขึ้นบันได และการเดินลงบันได โดยทำท่าแต่ละอิริยาบถประมาณ 3-4 นาที ตำแหน่งของสมาร์ทโฟน ได้แก่ กระเป๋ากางเกงขวา ดังภาพที่ 15 และเก็บข้อมูลจากเซนเซอร์ทั้งหมด 3 เซนเซอร์ได้แก่ เซนเซอร์วัดค่าความเร่ง เซนเซอร์วัดการหมุน และเซนเซอร์วัดสนามแม่เหล็กดังภาพที่ 16 ภายในชุดข้อมูลประกอบด้วยไฟล์ csv ทั้งหมด 6 ไฟล์ได้แก่ walking.csv,running.csv, sitting.csv, standing.csv, upstairs.csv และ downstairs.csv ตัวอย่างดังภาพที่ 17

ภาพที่ 16 ตำแหน่งของสมาร์ทโฟนที่กระเป๋ากางเกงขวา



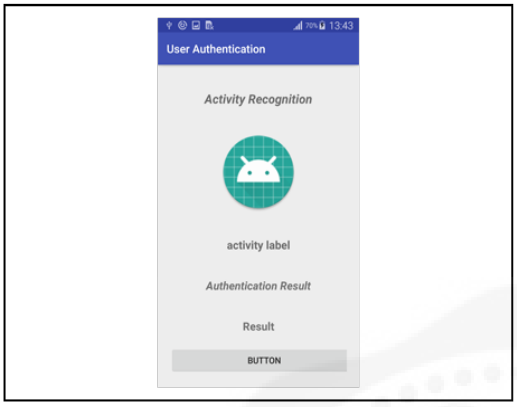
ภาพที่ 17 หน้าแอปพลิเคชันที่ใช้เก็บข้อมูลของเซนเซอร์

Ax	Ay	Az	Gx	Gy	Gz	Mx	My	Mz	Activity_Label
-5.405509	-6.353613	5.855619	-0.024501	-0.007457	-0.007457	-4.98	-19.799999	-57.419998	walking
-5.334281	-6.365584	6.029199	-0.024501	-0.007457	-0.007457	-4.80	-19.740000	-57.360001	walking
-5.233126	-6.370971	6.087856	-0.024501	-0.007457	-0.007457	-4.80	-19.620001	-57.299999	walking
-5.118205	-6.378153	6.021417	-0.024501	-0.007457	-0.007457	-4.80	-19.440001	-57.240002	walking
-4.992509	-6.407482	5.918467	-0.024501	-0.007457	-0.007457	-4.80	-19.320000	-57.180000	walking

ภาพที่ 18 ตัวอย่างข้อมูลเซนเซอร์ที่ได้จากอิริยาบถการเดิน

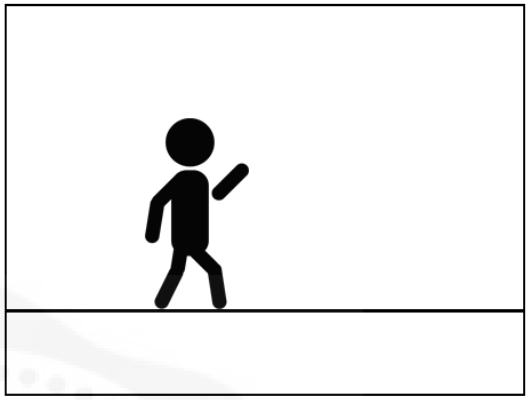
3.4.5 พัฒนาแอปพลิเคชันสำหรับยืนยันตัวตนผู้ใช้งาน

1



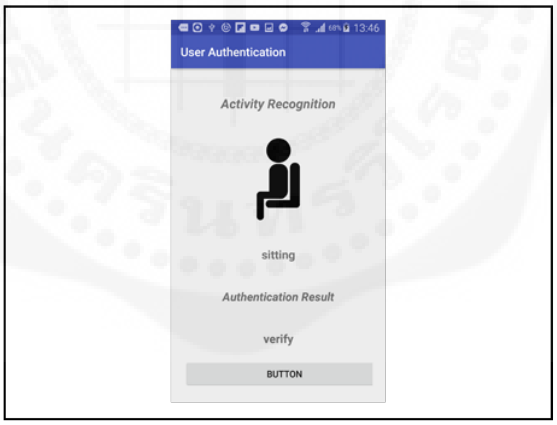
ทำการกดปุ่ม BUTTON บนแอปพลิเคชัน
เพื่อเริ่มทำการทดสอบการรู้จำอิริยาบถ

2



เริ่มทำการทดสอบ

3



หน้าแอปพลิเคชันจะแสดงให้เห็นว่ากำลังทำอิริยาบถใดอยู่

บทที่ 4
ผลการดำเนินงาน

4.1 ผลการประเมินประสิทธิภาพของการรู้จำอิริยาบถ

การรู้จำอิริยาบถใช้อัลกอริทึมในการจำแนกได้แก่อัลกอริทึมดังนี้

1. Support Vector Machines
2. k-Nearest Neighbors
3. Decision Tree
4. Bayesian networks (Gaussian Naive Bayes)

โดยผลการทดสอบของการรู้จำอิริยาบถแสดงด้วยค่า Accuracy

Algorithm	Position				Average Accuracy	Running time (second)
	pocket	arm	belt	wrist		
Decision Tree	**0.9846**	0.9844	0.9742	0.9534	0.97415	0.000997066
Support Vector Machine	0.9744	**0.9948**	0.9227	**0.9689**	0.9652	**0.000995028**
k-Nearest Neighbors	0.8615	0.9948	**0.9742**	0.8446	0.918775	0.021943331
Bayesian networks	0.9589	0.9430	0.8557	0.9637	0.930325	0.001024961

ตารางที่ 2 แสดงประสิทธิภาพของการรู้จำอิริยาบถ

4.2 ผลการประเมินประสิทธิภาพของการยืนยันตัวตน

การยืนยันตัวตนผู้ใช้งานเป็นปัญหาการจำแนกแบบคลาสเดียว(One-class classification) โดยวิธีการ ที่ใช้มีดังนี้

1. One-class SVM
2. One-Class K-Nearest Neighbor

โดยผลการทดสอบประสิทธิภาพของการยืนยันตัวตนผู้ใช้งาน ทดสอบด้วยค่า Accuracy, FRR (False Rejection Rate) และค่า FAR (False Acceptance Rate)

○ ไฟล์ pocket

Activity	Parameter	SVM	k-NN
walking	FRR	**13.25 %**	25.71 %
	FAR	**5.36 %**	11.49 %
	Accuracy	**94.39 %**	87.76 %
running	FRR	**15.11 %**	30.00 %
	FAR	**2.95 %**	9.11 %
	Accuracy	**96.52 %**	90.43 %
sitting	FRR	**19.25 %**	28.43 %
	FAR	**2.03 %**	2.73 %
	Accuracy	**97.35 %**	96.21 %
standing	FRR	**17.0 %**	28.57 %
	FAR	**3.65 %**	7.56 %
	Accuracy	**95.93 %**	91.34 %
upstairs	FRR	**15.50 %**	25.71 %
	FAR	21.42 %	**20.44 %**
	Accuracy	**82.64 %**	78.79 %
downstairs	FRR	**19.0 %**	30.00 %
	FAR	**17.90 %**	27.90 %
	Accuracy	**82.03 %**	71.40 %

ตารางที่ 3 ผลลัพธ์การยืนยันตัวตนของไฟล์ pocket

○ ไฟล์ arm

Activity	Parameter	SVM	k-NN
walking	FRR	**10.13 %**	24.29 %
	FAR	**12.41 %**	12.57 %
	Accuracy	**87.32 %**	86.52 %
running	FRR	**14.25 %**	20.00 %
	FAR	7.62 %	**6.00 %**
	Accuracy	91.87 %	**93.04 %**
sitting	FRR	**18.00 %**	31.43 %
	FAR	**3.33 %**	4.98 %
	Accuracy	**95.90 %**	94.01 %
standing	FRR	**17.0 %**	27.00 %
	FAR	**7.14 %**	15.49 %
	Accuracy	**92.36 %**	83.54 %
upstairs	FRR	**11.11 %**	25.71 %
	FAR	40.70 %	**23.30 %**
	Accuracy	60.18 %	**75.99 %**
downstairs	FRR	**19.50 %**	26.14 %
	FAR	27.55 %	**27.17 %**
	Accuracy	**72.67 %**	71.96 %

ตารางที่ 4 ผลลัพธ์การยืนยันตัวตนของไฟล์ arm

○ ไฟล์ wrist

Activity	Parameter	SVM	k-NN
walking	FRR	**16.00 %**	25.71 %
	FAR	**3.90 %**	4.92 %
	Accuracy	**95.46 %**	93.98 %
running	FRR	**17.25 %**	30.00 %
	FAR	8.15 %	**5.17 %**
	Accuracy	91.25 %	**93.58 %**
sitting	FRR	**16.6 %**	35.71 %
	FAR	4.0 %	**2.44 %**
	Accuracy	95.35 %	**96.40 %**
standing	FRR	**16.25 %**	32.86 %
	FAR	**3.14 %**	9.68 %
	Accuracy	**96.46 %**	89.16 %
upstairs	FRR	**19.25 %**	45.71 %
	FAR	5.52 %	**5.24 %**
	Accuracy	**94.05 %**	93.88 %
downstairs	FRR	**22.3 %**	35.71 %
	FAR	15.9 %	**7.84 %**
	Accuracy	**94.63 %**	90.90 %

ตารางที่ 5 ผลลัพธ์การยืนยันตัวตนของไฟล์ wrist

○ ไฟล์ belt

Activity	Parameter	SVM	k-NN
walking	FRR	**7.36 %**	28.57 %
	FAR	**13.20 %**	15.52 %
	Accuracy	**96.98 %**	83.54 %
running	FRR	**13.47 %**	32.86 %
	FAR	6.76 %	**4.48 %**
	Accuracy	92.43 %	**94.47 %**
sitting	FRR	**21.11 %**	31.43 %
	FAR	**3.65 %**	5.65 %
	Accuracy	**95.63 %**	93.14 %
standing	FRR	26.0 %	**25.71 %**
	FAR	14.35 %	**13.75 %**
	Accuracy	85.13 %	**87.12 %**
upstairs	FRR	**7.5 %**	27.14 %
	FAR	39.14 %	**35.27 %**
	Accuracy	61.82 %	**64.25 %**
downstairs	FRR	**23.61 %**	37.14 %
	FAR	**22.57 %**	39.56 %
	Accuracy	**77.39 %**	60.06 %

ตารางที่ 6 ผลลัพธ์การยืนยันตัวตนของไฟล์ belt

จากผลการทดสอบประสิทธิภาพของการยืนยันตัวตนผู้ใช้งานในแต่ละตำแหน่งของสมาร์ทโฟนและอิริยาบถ จึงได้ทำการเปรียบเทียบประสิทธิภาพของการยืนยันตัวตนในแต่ละไฟล์ ดังตารางที่ 6 จะเห็นได้ว่า ตำแหน่งกระเป๋ากางเกงด้านขวา อิริยาบถนั่งที่ถูกฝึกสอนด้วย One-class SVM มีค่าความถูกต้องมากที่สุดคือ 97.35 %

Activity	Performance Indicator	Algorithm							
		One-class SVM				One-class k-NN			
		pocket	arm	belt	wrist	pocket	arm	belt	wrist
walking	FRR	13.25 %	10.13 %	16.00 %	7.36 %	25.71 %	24.29 %	25.71 %	28.57 %
	FAR	5.36 %	12.41 %	3.90 %	13.20%	11.49 %	12.57 %	4.92 %	15.52 %
	Accuracy	94.39 %	87.32 %	95.46 %	86.98 %	87.76 %	86.52 %	93.98 %	83.54 %
running	FRR	15.11 %	14.25 %	17.25 %	13.47 %	30.00 %	20.00 %	30.00 %	32.86 %
	FAR	2.95 %	7.62 %	8.15 %	6.76 %	9.11 %	6.00 %	5.17 %	4.48 %
	Accuracy	96.52 %	91.87 %	91.25 %	92.43 %	90.43 %	93.04 %	93.58 %	94.47 %
sitting	FRR	19.25 %	18.00 %	16.60 %	21.11 %	28.43 %	31.43 %	35.71 %	41.43 %
	FAR	2.03 %	3.33 %	4.00 %	3.65 %	2.73 %	4.98 %	2.44 %	5.65 %
	Accuracy	97.35 %	95.90 %	95.35 %	95.63 %	96.21 %	94.01 %	96.40 %	93.14 %
standing	FRR	17.00 %	17.00 %	16.25 %	26.00 %	28.57 %	27.00 %	32.86 %	25.71 %
	FAR	3.65 %	7.14 %	3.14 %	14.35 %	7.56 %	15.49 %	9.68 %	13.75 %
	Accuracy	95.93 %	92.36 %	96.46 %	85.13 %	91.34 %	83.54 %	89.16 %	87.12 %
upstairs	FRR	15.50 %	11.11 %	19.25 %	7.50 %	25.71 %	25.71 %	45.71 %	27.14 %
	FAR	17.42 %	40.70 %	5.52 %	39.14 %	20.44 %	23.30 %	5.24 %	35.27 %
	Accuracy	82.64 %	60.18 %	94.05 %	61.82 %	78.79 %	75.99 %	93.88 %	64.25 %
downstairs	FRR	19.00 %	19.50 %	22.30 %	23.61 %	30.00 %	26.14 %	35.71 %	37.14 %
	FAR	17.90 %	27.55 %	19.90 %	22.57 %	27.90 %	27.17 %	7.84 %	39.56 %
	Accuracy	82.03 %	72.67 %	94.63 %	77.39 %	71.40 %	71.96 %	90.90 %	60.06 %

ตารางที่ 7 แสดงการเปรียบเทียบประสิทธิภาพของการยืนยันตัวตนแต่ละไฟล์

บทที่ 5
สรุปผลการดำเนินงาน

5.1 สรุปผลการดำเนินงาน

โครงงานการรู้จำอิริยาบถของมนุษย์โดยใช้ข้อมูลจากเซนเซอร์ของสมาร์ทโฟนสำหรับการยืนยันตัวตนผู้ใช้งานนี้ผู้จัดทำมีวัตถุประสงค์เพื่อพัฒนาแอปพลิเคชันที่มีการยืนยันตัวตนแบบพาสซีฟและต่อเนื่อง เพื่อ ลดความเสี่ยงต่อการถูกโจมตีจากความพยายามในการเข้าใช้งานสมาร์ทโฟน โดยใช้ลักษณะทางพฤติกรรม (behavioral biometrics)ในการตรวจสอบผู้ใช้งานสมาร์ทโฟนอย่างต่อเนื่อง โดยนำข้อมูลเหล่านี้มาวิเคราะห์ ด้วยวิธีการเรียนรู้ของเครื่อง (machine learning) ได้ โดยเราได้ทำการพัฒนาโมเดลขึ้นมา 2 โมเดล สำหรับ การพัฒนาแอปพลิเคชันคือโมเดลสำหรับการรู้จำอิริยาบถและโมเดลสำหรับการยืนยันตัวตนผู้ใช้งาน สมาร์ทโฟน โดยใช้ข้อมูลจากเซนเซอร์สมาร์ทโฟนที่เก็บจากการวางตำแหน่งสมาร์ทโฟน 4 ตำแหน่งที่ตัวผู้ใช้งานสมาร์ทโฟน ได้แก่ กระเป๋ากางเกงข้างขวา ข้อมือต้นแขน และเข็มขัด โมเดลที่ 1 : โมเดลสำหรับการรู้จำอิริยาบถ มี 3 ขั้นตอนหลัก ได้แก่ (1) การเตรียมข้อมูล (Data pre-processing) (2) การสกัดคุณลักษณะ (Feature extraction) (3) นำข้อมูลเข้าสู่กระบวนการฝึกสอนโดยใช้ SVM (Support Vector Machines), kNN (k-Nearest Neighbor), Decision Tree และ Bayesian networks เป็นอัลกอริทึมการเรียนรู้ของเครื่องที่ใช้ในการจำแนกอิริยาบถ โมเดลที่ 2 : โมเดลสำหรับการยืนยัน ตัวตนผู้ใช้งานสมาร์ทโฟน มี 3 ขั้นตอนหลัก ได้แก่ (1) การเตรียมข้อมูล (Data pre-processing) (2) การสกัดคุณลักษณะ (Feature extraction) (3) นำข้อมูลเข้าสู่กระบวนการฝึกสอนโดยใช้ One-class SVM และ One-class kNN ในการยืนยันตัวตนผู้ใช้งาน โดยสามารถสรุปผลได้ดังนี้

จากการทดสอบการรู้จำอิริยาบถ โมเดลจากการการฝึกสอนโดยใช้ SVM (Support Vector Machines) มีประสิทธิภาพการรู้จำอิริยาบถที่ดีที่สุดด้วยประสิทธิภาพ 99.48% บนบริเวณตำแหน่งต้นแขน ส่วนการยืนยันตัวตนผู้ใช้งานสมาร์ทโฟน โมเดลจากการฝึกสอนโดยใช้ one-class SVM ให้ค่าที่ดีที่สุดใน ตำแหน่งกระเป๋ากางเกงข้างขวา อิริยาบถการนั่ง 97.35% และได้นำโมเดลที่ได้มาพัฒนาเป็นแอปพลิเคชัน

จากผลการทดสอบแสดงให้เห็นว่า การยืนยันตัวตนผู้ใช้งานสมาร์ทโฟนยังให้ ประสิทธิภาพไม่ดีเท่าที่ควร เนื่องจากข้อมูลที่นำไปใช้ในการฝึกสอนนั้นมีข้อมูลบางส่วนที่ไม่สมบูรณ์ และ ไม่ถูกต้อง จึงทำให้อาจเกิดความผิดพลาดในการยืนยันตัวตนผู้ใช้งานสมาร์ทโฟนได้ ดังนั้นจึงควรฝึกสอน ข้อมูลของผู้ใช้งานซ้ำหลายๆ ครั้งเพื่อให้โมเดลการยืนยันตัวตนผู้ใช้งานสมาร์ทโฟนประสิทธิภาพมากยิ่งขึ้น

5.2 ปัญหาและอุปสรรคในการดำเนินงาน

1. โมเดลจาก One-class classifier นำมาใช้ในการพัฒนาแอปพลิเคชันสำหรับ Android ได้ยาก ซึ่งต้องมีการเปลี่ยนรูปแบบของข้อมูลของโมเดลก่อนที่จะนำมาใช้กับ Android เนื่องจากใช้รูปแบบไม่เหมือนกัน
2. ข้อมูลที่ใช้ในการฝึกสอนเก็บภายในระยะเวลา 3 นาที ทำให้อาจมีรูปแบบของอริยาบถที่ไม่ได้ทำ การเก็บข้อมูล

5.3 ข้อเสนอแนะ

1. ควรพัฒนารูปแบบการสร้างโมเดลให้มีการอัพเดตการฝึกสอนข้อมูลของผู้ใช้งานอยู่เป็นระยะ เพื่อให้ข้อมูลเป็นปัจจุบัน
2. แอปพลิเคชันการยืนยันตัวตนควรมีการตรวจหาตำแหน่งของสมาร์ทโฟน เพื่ออำนวยความสะดวกในการใช้งาน โดยที่ผู้ใช้ไม่ต้องทำการระบุว่ากำลังวางตำแหน่งของสมาร์ทโฟนไว้ส่วนไหนของร่างกาย ด้วยการใช้ classification position

บรรณานุกรม

[1] Physical activity recognition dataset, Sensors activity dataset. Retrieved December 21, 2018, from https://www.utwente.nl/en/eemcs/ps/research/dataset.

[2] L. Wang, T. Gu, X. Tao, and J. Lu, A hierarchical approach to real-time activity recognition in body sensor networks. (2012) Pervasive and Mobile Computing, 115- 130

[3] Thang Hoang, Deokjai Choi, และ Thuc Nguyen, Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme. (2015) International Journal of Information Security, 549–560

[4] Wei-Han Lee,Ruby B. Lee, Multi-sensor authentication to improve smartphone security. (2015) Information Systems Security and Privacy, 270-280

[5] Muhammad Muaaz, René Mayrhofer. Smartphone-Based Gait Recognition: From Authentication to Imitation. (2017) IEEE Transactions on Mobile Computing, 3209 - 3221

[6] Mohd Fikri Azli bin Abdullah, Ali Fahmi Perwira Negara, Md. Shohel Sayeed, Deok-Jai Choi and Kalaiarasi Sonai Muthu, Classification Algorithms in Human Activity Recognition using Smartphones. (2012) International Scholarly and Scientific Research & Innovation, 362-369

[7]

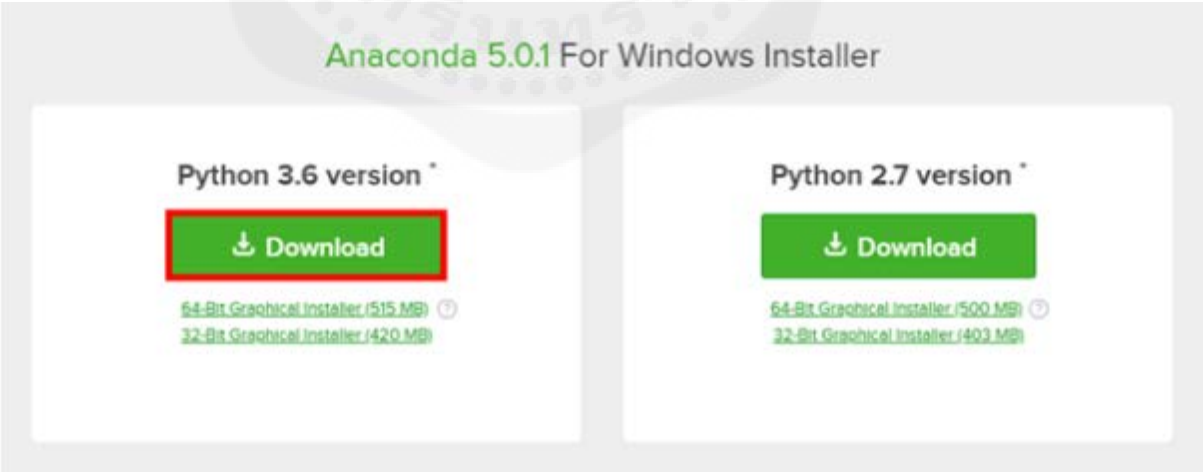
ภาคผนวก

ภาคผนวก ก
ขั้นตอนการติดตั้ง Anaconda

1. เข้าไปที่เว็บไซต์ https://www.anaconda.com/download/ เลือกดาวน์โหลดเวอร์ชันสำหรับระบบปฏิบัติการ Windows ดังภาพที่ ก-1

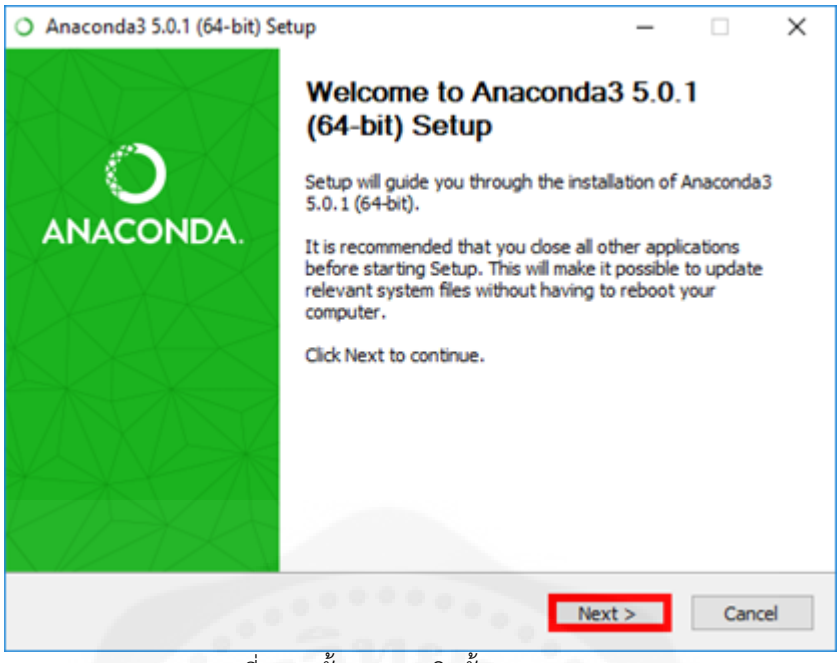
ภาพที่ ก-1 ขั้นตอนการติดตั้ง Anaconda (1)

2. คลิกที่ปุ่ม Download (Python 3.6 version) ดังรูป



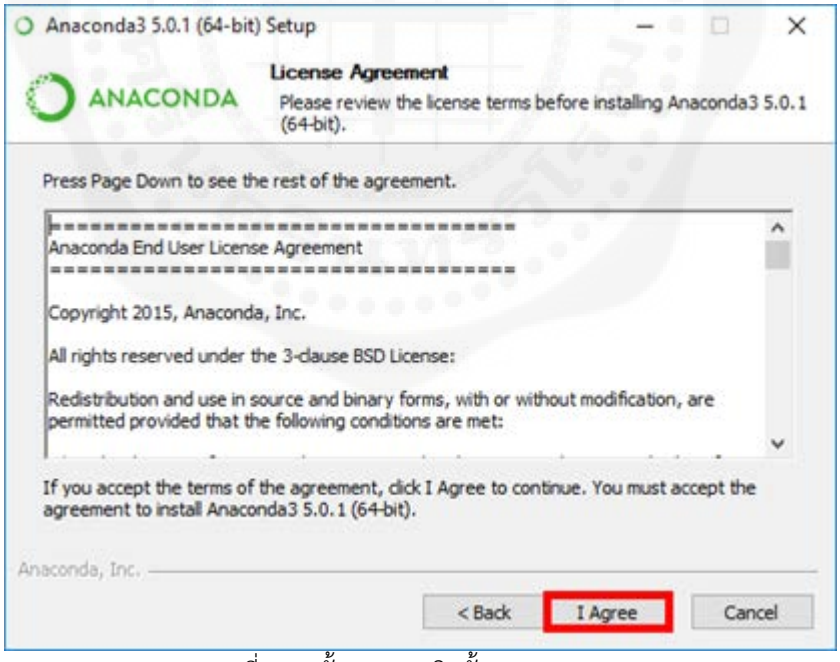
ภาพที่ ก-2 ขั้นตอนการติดตั้ง Anaconda (2)

3. ดับเบิลคลิกไฟล์ที่ดาวน์โหลดมาเพื่อติดตั้ง Anaconda คลิกที่ปุ่ม Next ดังภาพที่ ก-3



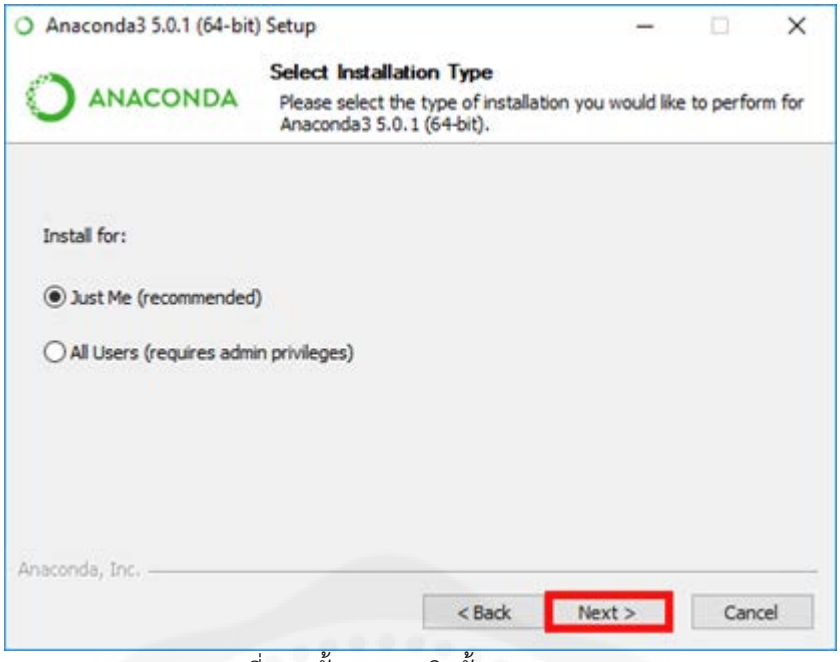
ภาพที่ ก-3 ขั้นตอนการติดตั้ง Anaconda (3)

4. คลิกที่ปุ่ม I Agree ดังภาพที่ ก-4



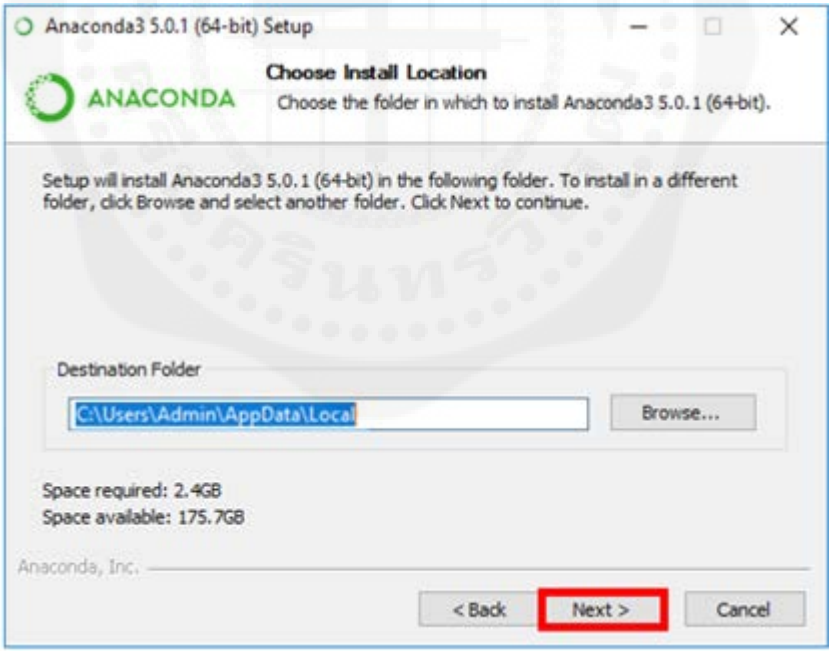
ภาพที่ ก-4 ขั้นตอนการติดตั้ง Anaconda (4)

5. เลือก Just Me (recommended) แล้วคลิกปุ่ม Next ดังภาพที่ ก-5



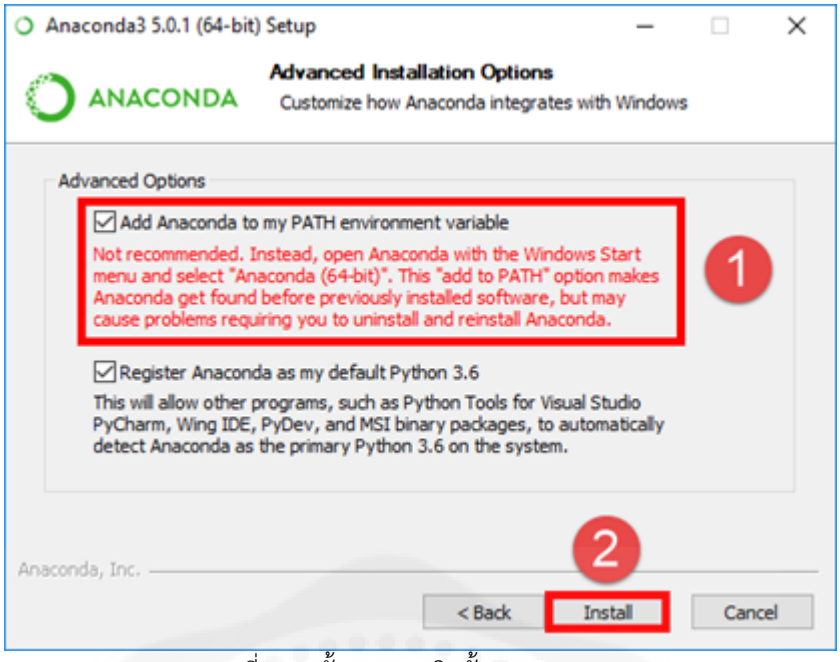
ภาพที่ ก-5 ขั้นตอนการติดตั้ง Anaconda (5)

6. คลิกที่ปุ่ม Next ดังภาพที่ ก-6



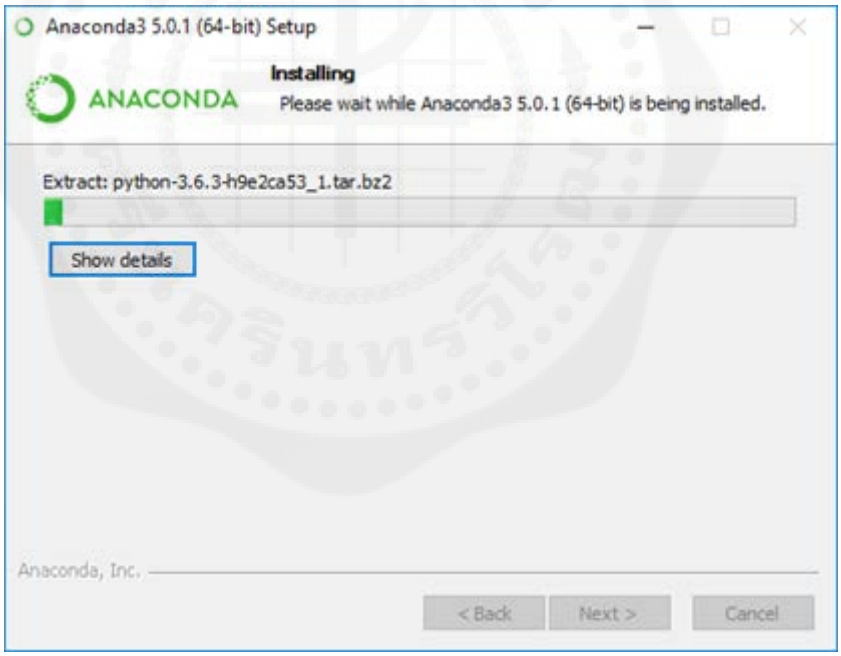
ภาพที่ ก-6 ขั้นตอนการติดตั้ง Anaconda (6)

7. เลือก Add Anaconda to my PATH environment variable แล้วคลิกปุ่ม Install ดังภาพที่ ก-7



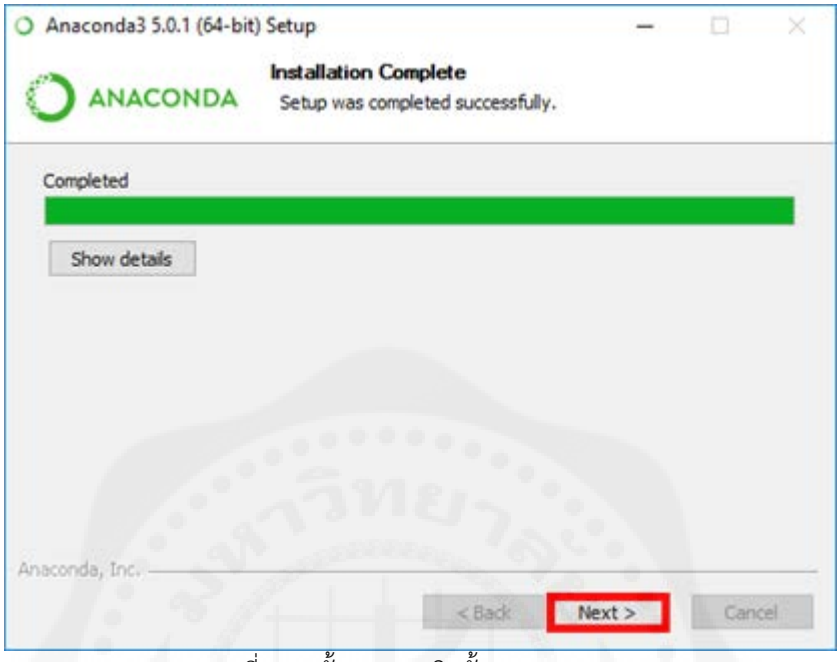
ภาพที่ ก-7 ขั้นตอนการติดตั้ง Anaconda (7)

8. รอจนกว่าการติดตั้งจะเสร็จสมบูรณ์ ดังภาพที่ ก-8



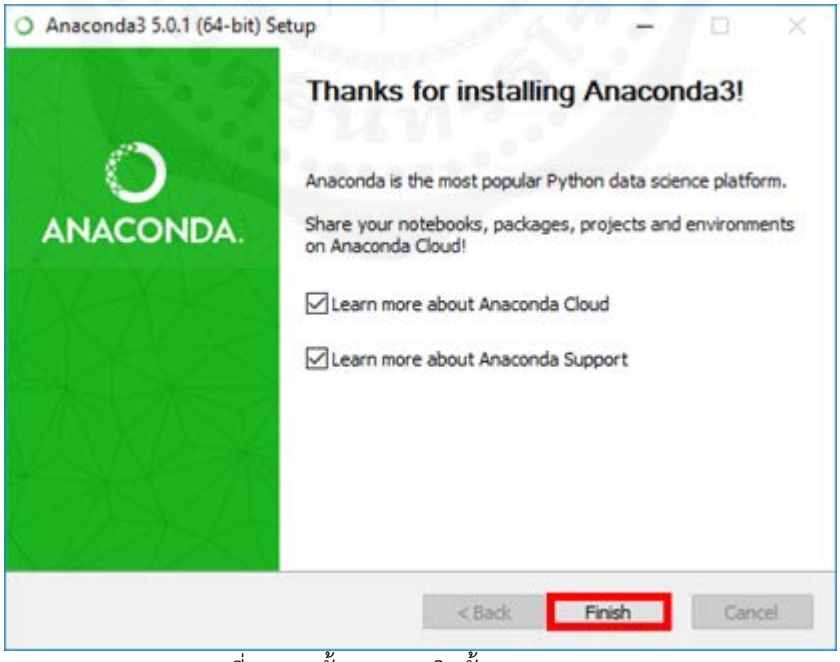
ภาพที่ ก-8 ขั้นตอนการติดตั้ง Anaconda (8)

9. เมื่อติดตั้งเสร็จสมบูรณ์แล้วให้คลิกปุ่ม Next ดังภาพที่ ก-9



ภาพที่ ก-9 ขั้นตอนการติดตั้ง Anaconda (9)

10. คลิกที่ปุ่ม Finish ดังภาพที่ ก-10



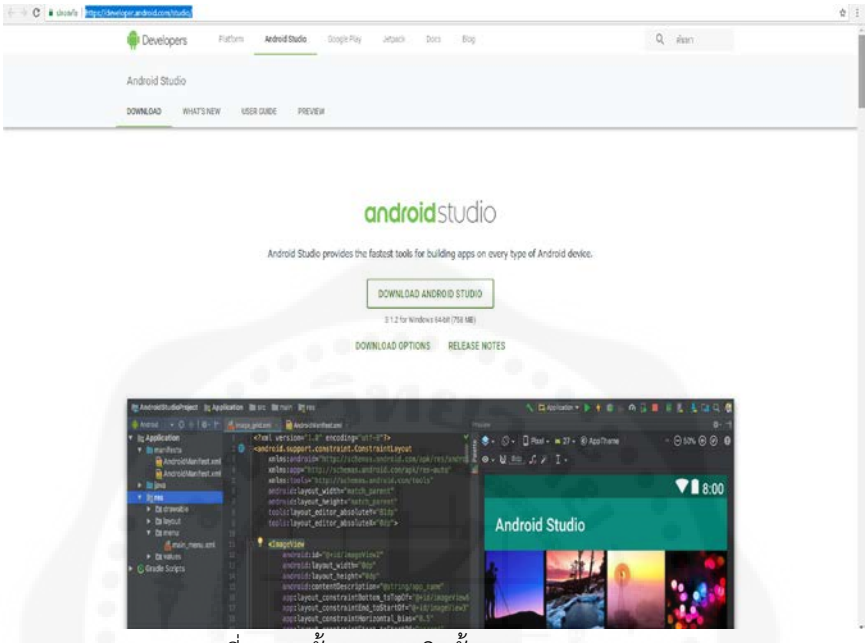
ภาพที่ ก-10 ขั้นตอนการติดตั้ง Anaconda (10)

11. เปิด Anaconda Navigotor ขึ้นมาและกด install Jupyter notebook ดังภาพที่ ก-11

ภาพที่ ก-11 ขั้นตอนการติดตั้ง Anaconda (11)

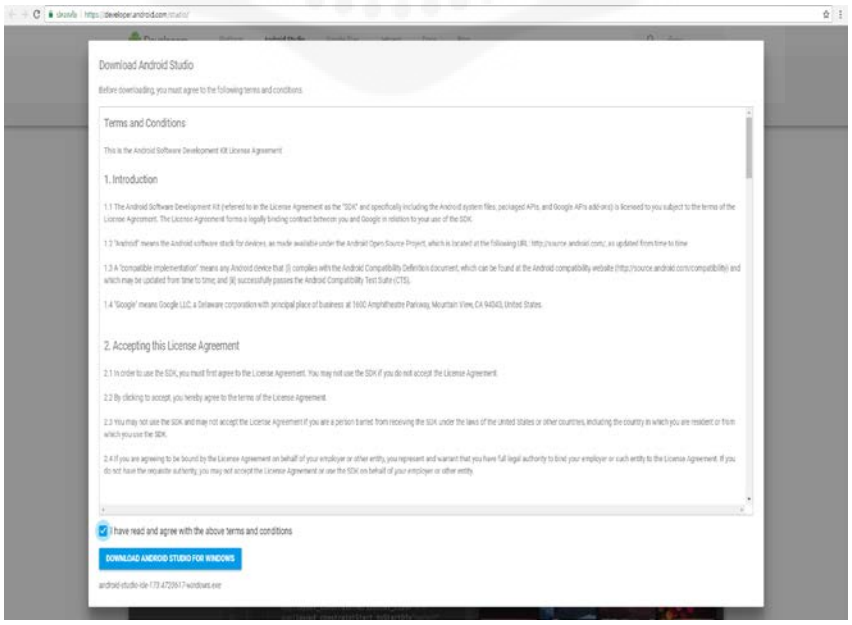
ภาคผนวก ข
ขั้นตอนการติดตั้ง Android Studio

1. เข้าเว็บไซต์ https://developer.android.com/studio/ เพื่อติดตั้ง Android Studio ดังภาพที่ ข-1



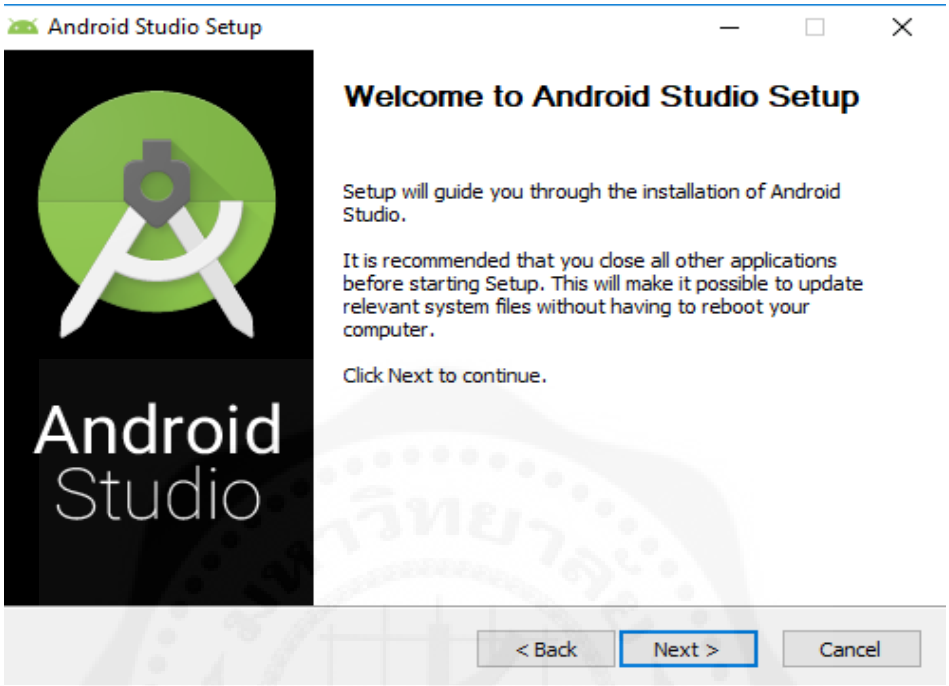
ภาพที่ ข-1 ขั้นตอนการติดตั้ง Android Studio (1)

2. เลือก I have read and agree with the above terms and conditions แล้วคลิก Download ดังภาพที่ ข-1 HYPERLINK "https://dl.google.com/dl/android/studio/install/3.1.2.0/android-studio-ide-173.4720617-windows.exe"



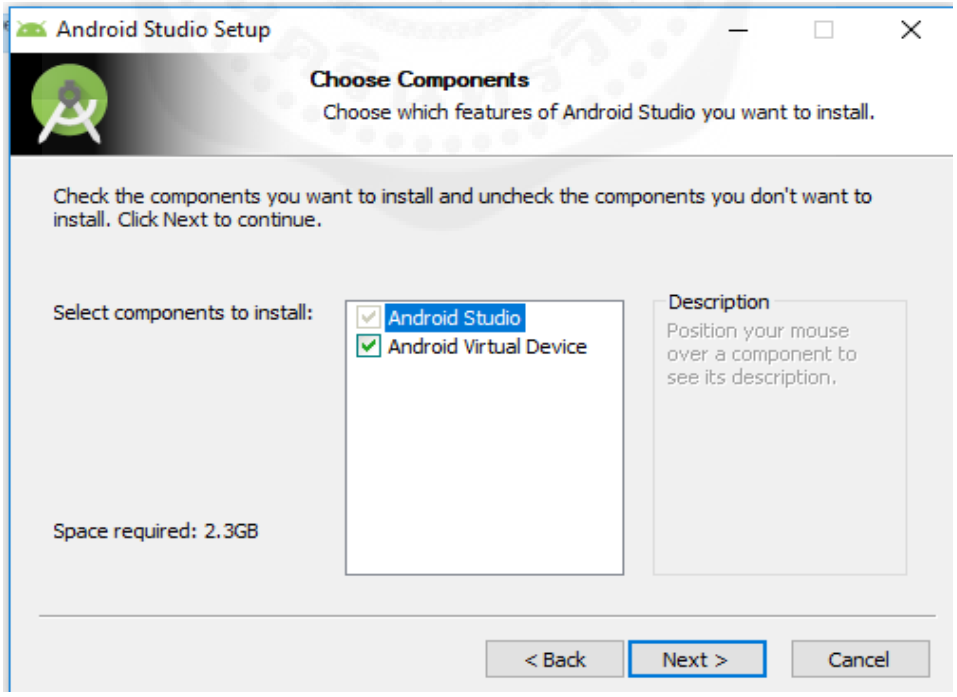
ภาพที่ ข-2 ขั้นตอนการติดตั้ง Android Studio (2)

3. เปิดไฟล์ที่ดาวน์โหลดมา หลังจากนั้นจะปรากฏหน้าจอการติดตั้ง คลิก next ดังภาพที่ ข-3



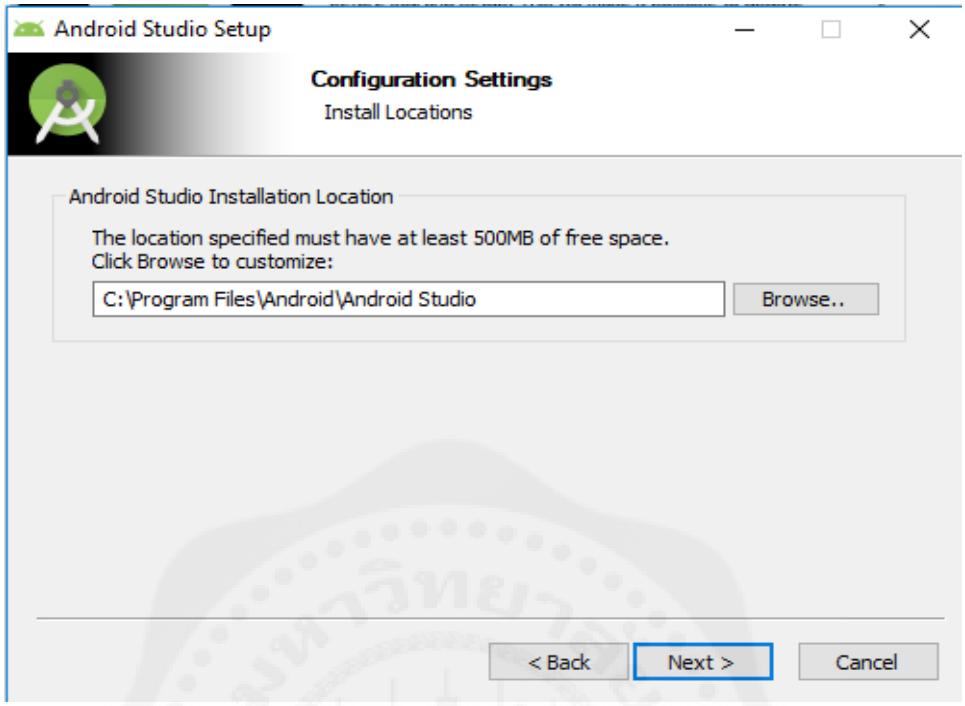
ภาพที่ ข-3 ขั้นตอนการติดตั้ง Android Studio (3)

4. คลิกที่ปุ่ม next ดังภาพที่ ข-4



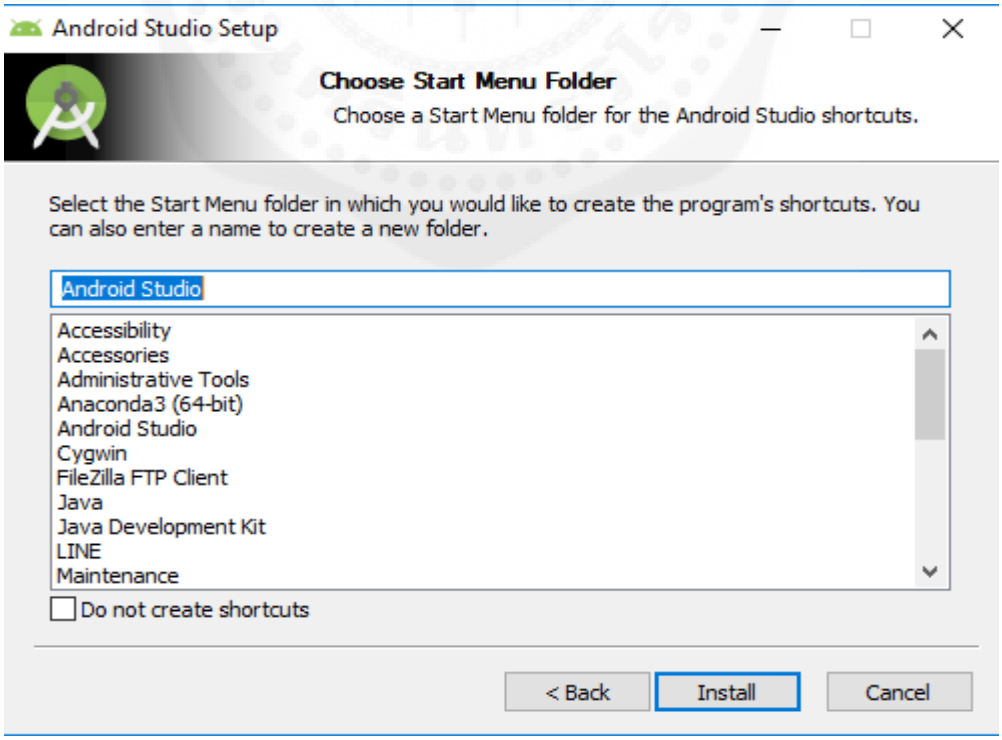
ภาพที่ ข-4 ขั้นตอนการติดตั้ง Android Studio (4)

5. คลิกที่ปุ่ม next ดังภาพที่ ข-5



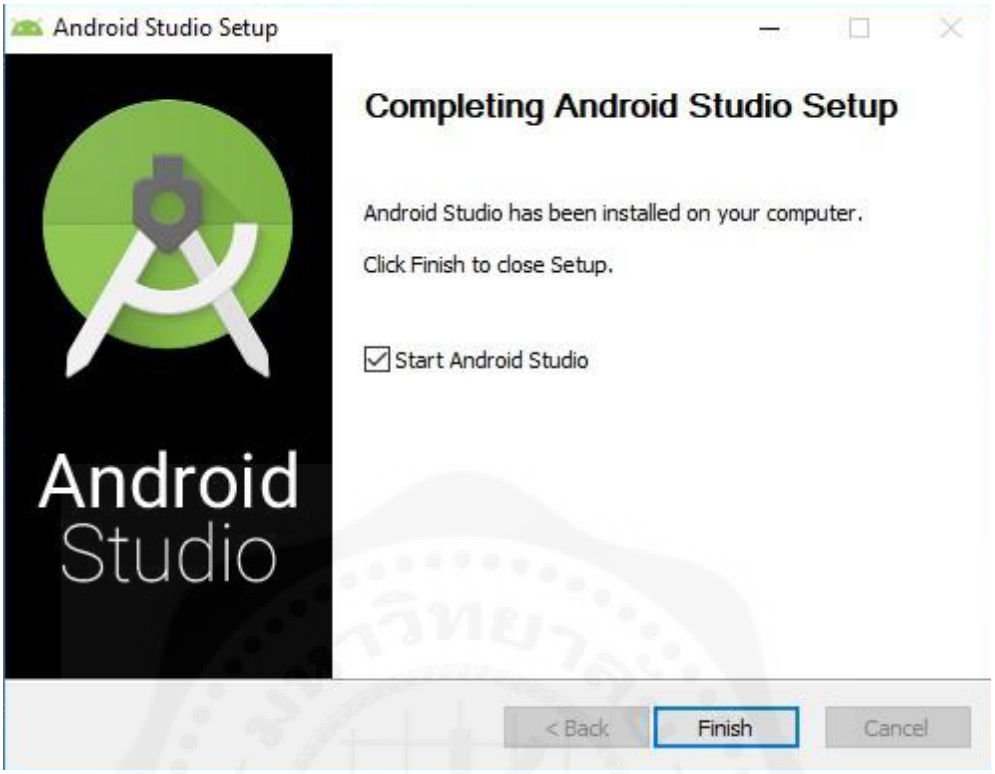
ภาพที่ ข-5 ขั้นตอนการติดตั้ง Android Studio (5)

6. หากไม่ต้องการให้สร้าง shortcuts ที่ desktop ให้เลือกที่ Do not create shortcuts แล้วคลิก install ดังภาพที่ ข-6 และรอจนโปรแกรมเสร็จสิ้น



ภาพที่ ข-6 ขั้นตอนการติดตั้ง Android Studio (6)

7. เสร็จสิ้นการติดตั้งโปรแกรม คลิกที่ปุ่ม Finish ดังภาพที่ ข□7



ภาพที่ ข-7 ขั้นตอนการติดตั้ง Android Studio (7)

ภาคผนวก ค

วิธีการใช้แอปพลิเคชันในการยืนยันตัวตน