Trajectory Data Mining Based on Location History

การทำเหมืองข้อมูลเส้นทางเดินจากข้อมูลการระบุตำแหน่ง

พิชญา ลีฬหาทร
Pichaya Leerahathorn

ภัทร โตจิตต์
Pat Tochit

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2561

Trajectory Data Mining Based on Location History

การทำเหมืองข้อมูลเส้นทางเดินจากข้อมูลการระบุตำแหน่ง

พิชญา ลีฬหาทร
Pichaya Leerahathorn

ภัทร โตจิตต์
Pat Tochit

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2561

คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

โครงงาน	Trajectory Data Mining Based on Location History การทำเหมืองข้อมูลเส้นทางเดินจากข้อมูลการระบุตำแหน่ง
นิสิต	นางสาวพิชญา ลีฬหาทร 58102010814 นางสาวภัทร โตจิตต์ 58102010819
ปริญญา	วิทยาศาสตรบัณฑิต (วท.บ.)
ภาควิชา	วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาโครงงาน	ดร.วีรยุทธ เจริญเรืองกิจ

ลงชื่อ.....

(ดร.วีรยุทธ เจริญเรืองกิจ)

บทคัดย่อ

ข้อมูลการเคลื่อนที่ของผู้คนสามารถแสดงข้อมูลที่ซ่อนอยู่มากมายเกี่ยวกับพฤติกรรมของผู้คนและพื้นที่ที่พวกเขาสนใจ ซึ่งสามารถแบ่งออกเป็นสองปัญหาย่อย โดยปัญหาย่อยแรกคือการกำหนดตำแหน่งของแต่ละบุคคลตามข้อมูลด้วยพิกัด GPS ปัญหาย่อยที่สองคือการดึงข้อมูลที่เป็นประโยชน์จากลำดับการระบุตำแหน่งจำนวนมากที่ได้รับจากปัญหาย่อยแรก ในปัญหาย่อยแรกนั้น การทดลองภายนอกอาคารนั้นได้มีการทำวิจัยอย่างมีประสิทธิภาพโดยการประมาณตำแหน่งจากข้อมูล GPS ในขณะที่การทดลองภายในอาคารยังคงมีการทำการวิจัยอยู่อย่างต่อเนื่องโดยใช้เซ็นเซอร์หลายประเภท ปัญหาย่อยที่สองมักถูกจัดเป็นการทำเหมืองข้อมูลเส้นทางเดิน ซึ่งยังคงเป็นที่สนใจ งานวิจัยนี้มีความสนใจในการทำเหมืองข้อมูลเส้นทางเดินภายในอาคาร แต่เนื่องจากข้อจำกัดของเวลาและทรัพยากร ทำให้งานวิจัยนี้ดำเนินการบนชุดข้อมูลภายนอกอาคาร ในขณะที่ปัญหาย่อยแรกจะถูกรวบรวมจากบีคอนส์ 3 ตัว ภายในอาคารวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒเพื่อกำหนดวิธีที่ดีที่สุดในการระบุตำแหน่งในอาคาร ในการทำนายตำแหน่งใช้อัลกอริทึม K-Nearest Neighbor, Random Forest และ Support Vector Classification ซึ่ง Support Vector Classification ได้ค่าความคลาดเคลื่อนโดยเฉลี่ยเป็น 4.45 เมตรและปัญหาการทำเหมืองข้อมูลเส้นทางในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อค้นหาตำแหน่งที่ได้รับความสนใจของแต่ละผู้ใช้ด้วยตำแหน่งพิกัดจากชุดข้อมูล Geolife GPS trajectory ที่ดำเนินการโดยวิธี DBSCAN

Abstract

People movement data can reveal much hidden information about people's behavior and their area of interest. The problem can be divided into two sub-problems. The first sub-problem is to determine the location of individual based on the sensing data such as GPS coordinates. The second sub-problem is to extract useful information from a large amount of sequential location data obtained from the first sub-problem. The first sub-problem for the outdoor setting has been well-investigated by estimating the location from the GPS data while the in-door setting is still an ongoing research using several types of sensors. The second sub-problem is often categorized as trajectory mining, which is still open for many research questions. This research is interested in indoor trajectory mining. Due to the limitation of time and resources, the trajectory mining problem is performed on the outdoor dataset while the first sub-problem are collected from 3 beacons inside the Science building, Srinakharinwirot University, to determine the best algorithms for the indoor positioning. The positioning prediction obtained K-Nearest Neighbor, Random Forest, Support Vector Classification 4.45 meters in error on average. The trajectory mining problem in this research aims to find the area of interest of people given a set of coordinate positions from the Geolife GPS trajectory dataset implemented by the DBSCAN method.

กิตติกรรมประกาศ

การทำเหมืองข้อมูลเส้นทางเดินจากข้อมูลการระบุตำแหน่ง (Trajectory Data Mining Based on Location History) สำเร็จลุล่วงได้ด้วยความกรุณาช่วยเหลือ ให้คำแนะนำและคำปรึกษาจาก อ.ดร.วีรยุทธ เจริญเรืองกิจ อาจารย์ที่ปรึกษาของโครงงาน ที่กรุณาเสียสละเวลาเพื่อให้คำปรึกษา แนะนำแนวทางในการแก้ปัญหาและแก้ไขข้อบกพร่องที่เกิดขึ้นระหว่างการทำงาน คณะผู้จัดทำกราบขอบพระคุณเป็นอย่างสูง

ขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์ที่สนับสนุนอุปกรณ์สถานที่ในการทำโครงงาน และ ขอขอบพระคุณอาจารย์ทุกๆ ท่านในภาควิชาวิทยาการคอมพิวเตอร์ที่กรุณาให้ความรู้ ทำให้คณะผู้จัดทำสามารถนำความรู้ที่ได้มาประยุกต์ใช้ในการดำเนินโครงงาน รวมทั้งให้คำปรึกษาและคำแนะนำที่เป็นประโยชน์ ทำให้คณะผู้จัดทำสามารถแก้ไขโครงงานได้สมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณมหาวิทยาลัยศรีนครินทรวิโรฒที่เป็นแหล่งศึกษาหาความรู้ และเอื้อเฟื้อสถานที่สำหรับการดำเนินโครงงาน รวมทั้งให้ประสบการณ์ในการทำกิจกรรมต่างๆ เพื่อที่จะนำไปประยุกต์ใช้ในการทำงานและการดำเนินชีวิตต่อไป

ขอขอบพระคุณบิดา มารดา เพื่อน รวมถึงครอบครัวของคณะผู้จัดทำที่ได้ให้การช่วยเหลือและสนับสนุนมาโดยตลอด รวมถึงให้ความห่วงใยและเป็นกำลังใจสำคัญในการทำโครงงานเสมอมา

สุดท้ายขอขอบพระคุณผู้มีส่วนเกี่ยวข้องทุกๆ ท่านที่ทำให้โครงงานนี้สำเร็จลุล่วงไปได้ด้วยดี คณะผู้จัดทำหวังว่าโครงงานการทำเหมืองข้อมูลเส้นทางเดินจากข้อมูลการระบุตำแหน่ง (Trajectory Data Mining Based on Location History) จะเป็นประโยชน์ต่อผู้ที่สนใจศึกษา และสามารถเป็นส่วนหนึ่งสำหรับผู้ที่สนใจจะนำไปต่อยอดในบริบทต่างๆ ต่อไป

คณะผู้จัดทำ

สารบัญ

บทคัดย่อ ..... ก
Abstract ..... ข
กิตติกรรมประกาศ ..... ค
สารบัญ ..... ง
สารบัญภาพ ..... ฉ
สารบัญตาราง ..... ช
บทที่ 1 1
1.1 ที่มาและความสำคัญ 1
1.2 วัตถุประสงค์ของงานวิจัย 1
1.3 ขอบเขตของงานวิจัย 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ 2
บทที่ 2 3
2.1 ทฤษฎีและแนวคิดที่ใช้ในงานวิจัย 3
2.1.1 ระบบระบุตำแหน่งภายในอาคาร (Indoor Positioning System: IPS) 3
2.1.2 Machine learning 4
2.1.3 เทคนิค Classification 5
2.1.4 เทคนิคการวิเคราะห์กลุ่ม (Clustering analysis) 7
2.2 งานวิจัยที่เกี่ยวข้อง 8
2.2.1 งานวิจัยเรื่อง“Trajectory Clustering via Deep Representation Learning” 8
2.2.2 งานวิจัยเรื่อง “Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web” 8
2.2.3 งานวิจัยเรื่อง “Smartphone Inertial Sensor-Based Indoor Localization and Tracking With iBeacon Corrections” 9
2.2.4 งานวิจัยเรื่อง “Mining Individual Life Pattern Based on Location History” 9
2.2.5 งานวิจัยเรื่อง “Beacon Placement for Range-Based Indoor Localization” 9

บทที่ 3 11
3.1 แผนการดำเนินงาน 11
3.2 แผนการดำเนินงานตลอดโครงงาน 13
3.3 อุปกรณ์และเครื่องมือที่ใช้ 14
3.3.1 ฮาร์ดแวร์ (Hardware) 14
3.3.2 ซอร์ฟแวร์ (Software) 14
3.3.3 ภาษาที่ใช้ (Language) 14
บทที่ 4 15
4.1ผลลัพธ์จากการเก็บข้อมูล 15
4.2 ผลลัพธ์จากการคำนวณข้อมูล 17
4.3 ผลลัพธ์ที่ได้จากการทดสอบ 18
4.4 ผลลัพธ์สำรวจและทำการเตรียมข้อมูล 18
4.5 ผลลัพธ์การหาตำแหน่งหยุด (Stay point) ของแต่ละผู้ใช้ 21
4.6 ผลลัพธ์การนำตำแหน่งหยุดมาจัดกลุ่มโดยใช้ DBSCAN 22
4.7 ผลลัพธ์การนำตำแหน่งที่ได้รับความสนใจมาประเมินความถูกต้อง ... 23
บทที่ 5 24
5.1 สรุปผลการดำเนินโครงงาน 24
5.2 ปัญหาและอุปสรรคในการดำเนินโครงงาน 24
5.3 ข้อเสนอแนะ 25
บรรณานุกรม 26
ภาคผนวก 29
ภาคผนวก ก 30
ภาคผนวก ข 34
ภาคผนวก ค 37
ภาคผนวก ง 38

สารบัญภาพ

ภาพที่ 1 ตัวอย่าง Support Vector Classification 6
ภาพที่ 2 แผนการดำเนินงานโดยรวม 11
ภาพที่ 3 แผนผังที่ทำการเก็บข้อมูลบริเวณชั้น 18..... 15
ภาพที่ 4 ตัวอย่างข้อมูลที่เก็บได้..... 16
ภาพที่ 5 ผลการคำนวณหาตำแหน่ง 17
ภาพที่ 6 ผลลัพธ์การทำนายข้อมูลแบบเส้นทางเดินโดยใช้อัลกอริทึมแต่ละแบบ 17
ภาพที่ 7 ตัวอย่างข้อมูลเส้นทางเดินของผู้ใช้ 19
ภาพที่ 8 ตำแหน่งหยุด 21
ภาพที่ 9 ผลลัพธ์ของผู้ใช้ที่ 1 22

สารบัญตาราง

ตารางที่ 1 ผลการทดสอบความแม่นยำและความคลาดเคลื่อน 18
ตารางที่ 2 แสดงรายละเอียดชุดข้อมูล Geolife GPS trajectory 19
ตารางที่ 3 แสดงรายละเอียดชุดข้อมูล Geolife GPS trajectory ที่นำมาใช้ 20
ตารางที่ 4 แสดงรายละเอียดตำแหน่งหยุด 21
ตารางที่ 5 แสดงจำนวนตำแหน่งที่ได้รับความสนใจของแต่ละผู้ใช้ 22
ตารางที่ 6 แสดงจำนวนตำแหน่งที่ได้รับความสนใจที่ถูกต้องของแต่ละผู้ใช้ 23

บทที่ 1
บทนำ

1.1ที่มาและความสำคัญ

เนื่องจากในปัจจุบันมีการใช้เทคโนโลยีอย่างแพร่หลายจึงทำให้เกิดข้อมูลอิเล็กทรอนิกส์จำนวนมาก ไม่ว่าจะได้มาจากการใช้ไวไฟ การใช้บลูทูธ หรือ การใช้ระบบระบุตำแหน่งบนพื้นโลก (Global Positioning System: GPS) ในสถานการณ์ที่หลากหลายไม่ว่าจะเป็นการระบุตำแหน่งของวัตถุ (Positioning) หรือติดตามวัตถุ (Tracking) หรือการใช้โทรศัพท์เชื่อมต่อระบบเครือข่ายอินเทอร์เน็ตเพื่อใช้ในการสื่อสารหรือค้นหาข้อมูลต่างๆ เป็นต้น ซึ่งข้อมูลเหล่านี้อาจให้ประโยชน์ในการพัฒนาในด้านต่างๆ เช่น การจัดการด้านการคมนาคม

ข้อมูลต่างๆที่เกิดขึ้นนั้นสามารถนำไปประยุกต์เข้ากับการเรียนรู้ของเครื่อง (Machine Learning) โดยการเรียนรู้ของเครื่องจะเข้ามามีบทบาทในการวิเคราะห์ข้อมูลที่มีจำนวนมากจากการเรียนรู้ด้วยตนเองและแสดงผลลัพธ์ที่น่าพึงพอใจ ซึ่งผลลัพธ์ที่ได้ออกมานั้นสามารถกำหนดได้ตามที่ต้องการ ยกตัวอย่างเช่น การนำข้อมูลจากระบบระบุตำแหน่งบนพื้นโลกในการเดินทาง เช่น ละติจูด (Latitude) ลองจิจูด (Longitude) เวลา (Time stamp) ไปทำเหมืองข้อมูล (Data Mining) ซึ่งสามารถนำข้อมูลไปหาตำแหน่งที่ได้รับความสนใจได้อีกด้วย

ทางผู้วิจัยได้เล็งเห็นว่าการจะระบุตำแหน่งภายในอาคารนั้นมีความไม่แม่นยำเมื่อระบุตำแหน่งด้วยระบบระบุตำแหน่งบนพื้นโลก จึงได้ทำการศึกษาระบบระบุตำแหน่งภายในอาคารโดยใช้การระบุตำแหน่งจากบีคอนส์และเรียนรู้การเก็บข้อมูล และได้ทำการศึกษา ทำความเข้าใจเกี่ยวกับการค้นหาตำแหน่งที่ได้รับความสนใจจากการเดินทาง เพื่อนำไปประยุกต์กับการหาตำแหน่งที่ได้รับความสนใจในบริเวณต่างๆและนำข้อมูลใช้ในการวางแผนการตลาดได้

1.2วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาวิธีการเก็บข้อมูลการระบุตำแหน่งด้วยการวัดค่า RSSI จากบีคอนส์และนำไปวิเคราะห์เพื่อระบุตำแหน่ง
2. เพื่อศึกษาอัลกอริทึมต่างๆ ที่เกี่ยวข้องกับข้อมูลการเดินทาง
3. เพื่อศึกษาการค้นหาตำแหน่งที่ได้รับความสนใจ

1.3ขอบเขตของงานวิจัย

1. เรียนรู้วิธีการเก็บข้อมูลการระบุตำแหน่งจากบีคอนส์และทดลองการระบุตำแหน่ง โดยการทดลองจัดทำที่ ชั้น 18 อาคาร 19 คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
2. ศึกษาการใช้ Machine Learning เพื่อระบุตำแหน่งที่ได้รับความสนใจ (point of interested) โดยใช้ชุดข้อมูล Geolife GPS trajectory

1.4ประโยชน์ที่คาดว่าจะได้รับ

1. เข้าใจพฤติกรรมการเดินทางจากข้อมูลจำนวนมาก
2. นำไปประยุกต์ใช้ในการจัดตำแหน่งของร้านค้าภายในห้างสรรพสิ่งสินค้า
3. สามารถนำไปสู่การแนะนำสถานที่ที่เป็นตำแหน่งที่ได้รับความสนใจ
4. เข้าใจอัลกอริทึมต่างๆเกี่ยวกับข้อมูลการเดินทางและการทำงานกับข้อมูลจำนวนมาก

บทที่ 2
องค์ความรู้ที่เกี่ยวข้อง

2.1 ทฤษฎีและแนวคิดที่ใช้ในงานวิจัย

2.1.1 ระบบระบุตำแหน่งภายในอาคาร (Indoor Positioning System: IPS) [1]

IPS เป็นระบบที่ถูกใช้เพื่อระบุตำแหน่งของสิ่งของหรือบุคคลภายในอาคารโดยใช้แสง คลื่นวิทยุ สนามแม่เหล็ก สัญญาณเสียงหรือข้อมูลจากตัวรับสัญญาณต่างๆ ซึ่ง IPS สามารถใช้เทคโนโลยีที่แตกต่างกันรวมถึงการวัดระยะทางไปยังโหนดที่มีตำแหน่งคงที่ เช่น Wi-Fi, Bluetooth beacons, Magnatic positioning

2.1.1.1 Bluetooth [2]

ใช้การกวาดหาสัญญาณด้วยบลูทูธ โดยใช้เทคโนโลยีที่เรียกว่า Bluetooth Low Energy หรือ BLE ซึ่งต่างจากบลูทูธปกติที่ใช้จับคู่อุปกรณ์พกพาที่มักใช้กันบ่อยๆ ซึ่งการใช้การกวาดจับสัญญาณด้วย BLE นี้ จำเป็นที่เจ้าของอุปกรณ์ลูกต้องรับรู้และกดอนุญาตด้วย BLE นี้ใช้พลังงานได้ประหยัดกว่าบลูทูธทั่วไปมาก แบตเตอรี่ชุดหนึ่งสามารถทำงานต่อเนื่องได้หลายปีโดยไม่ต้องชาร์จใหม่ แต่ก็มีข้อจำกัดตามธรรมชาติของบลูทูธคือ เหมาะกับการรับส่งสัญญาณในระยะใกล้ๆ เท่านั้น ตัวกวาดสัญญาณจับบลูทูธ หรือที่เรามักเรียกว่า บีคอนส์ หรือ Active RFID นี้ จะส่งสัญญาณกวาดรอบตัวเหมือนประภาคาร ซึ่งตัวรับสัญญาณอย่างสมาร์ทโฟน จะเอาสัญญาณบลูทูธที่ได้มาประมวลผล ทำงานร่วมกับแอปพลิเคชันหรือเครือข่ายอื่นสำหรับตีความตำแหน่งที่ตั้งโดยเฉพาะ

2.1.1.2 การวัดค่าความเข้มของสัญญาณ (RSSI) [3]

RSSI (Receive Signal Strength Indicator) เป็นวิธีการวัดความเข้มสัญญาณของสัญญาณวิทยุ ระหว่างเซ็นเซอร์ไร้สายตัวหนึ่งไปยังเซ็นเซอร์ไร้สายอีกตัวหนึ่ง โดยมีความสัมพันธ์ตามสมการดังนี้

$$RSSI = -(10nlog_{10}d + A)$$

n คือค่าคงที่ของการกระจายสัญญาณ

d คือระยะห่างระหว่างเซ็นเซอร์แต่ละตัว

A คือค่าสัมบูรณ์ความเข้มสัญญาณที่ได้รับในระยะ 1 เมตร

ค่า RSSI นั้นมีหน่วยเป็นเดซิเบลเมตร (dBm) เมื่อวิเคราะห์สูตรจะพบว่าสามารถใช้ความสัมพันธ์ระหว่างความเข้มของสัญญาณกับ

ระยะทางได้ เมื่อ d มีค่ามากส่งผลให้ค่าความเข้มของสัญญาณมีค่าลดน้อยลง

2.1.2 Machine learning [4]

การเรียนรู้ของเครื่อง (Machine Learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจากการศึกษาการรู้จำแบบ เกี่ยวข้องกับการศึกษาและการสร้างอัลกอริทึมที่สามารถเรียนรู้ข้อมูลและทำนายข้อมูลได้ อัลกอริทึมนั้นจะทำงานโดยอาศัยโมเดลที่สร้างมาจากชุดข้อมูลตัวอย่างขาเข้าเพื่อการทำนายหรือตัดสินใจในภายหลัง แทนที่จะทำงานตามลำดับของคำสั่งโปรแกรมคอมพิวเตอร์

การเรียนรู้ของเครื่อง (Machine Learning) สามารถสร้างโมเดลได้หลักๆมี 3 ประเภทดังนี้

(1) การเรียนรู้แบบมีผู้สอน (Supervised Learning)

Supervised Learning คือ การที่โปรแกรมจะสามารถจำแนกได้ว่าข้อมูลที่ใส่เข้ามาหมายถึงอะไร เป็นเทคนิคการเรียนรู้การฝึกฝนจากข้อมูลอินพุต (Input) โดยที่มีคำตอบที่ถูกต้อง (Label) ของข้อมูลเอาต์พุต (Output) อยู่แล้ว โดยผลการเรียนรู้จะเป็นได้ทั้งฟังก์ชันที่ให้ค่าต่อเนื่อง (Regression) หรือฟังก์ชันที่ใช้ทำนายวัตถุ (Classification)

(2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

Unsupervised Learning เป็นเทคนิคการเรียนรู้ที่มีข้อมูลอินพุต (Input) แต่ที่ไม่มีคำตอบที่ถูกต้อง (Label) ของข้อมูลเอาต์พุต (Output)

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) สามารถนำไปใช้ร่วมกับทฤษฎีของเบย์ (Bayes' Theorem) เพื่อหาความน่าจะเป็นแบบมีเงื่อนไขของตัวแปรสุ่มโดยกำหนดตัวแปรที่เกี่ยวข้องให้ นอกจากนี้ยังสามารถนำไปใช้ในการบีบอัดข้อมูล ซึ่งโดยพื้นฐานแล้ว ขั้นตอนวิธีการบีบอัดข้อมูลจะขึ้นอยู่กับการแจกแจงความน่าจะเป็นของข้อมูล

(3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)

Reinforcement learning เป็นเทคนิคการเรียนรู้ที่คอมพิวเตอร์มีปฏิสัมพันธ์กับสิ่งแวดล้อมที่เปลี่ยนไปตลอดเวลา โดยคอมพิวเตอร์จะต้องทำงานบางอย่าง โดยที่ไม่รู้วิธีการที่ทำอยู่นั้นเข้าใกล้เป้าหมายแล้วหรือไม่ ตัวอย่างเช่น การเล่นหมากรุก

จะต้องมีการทำนายล่วงหน้าว่าจะสามารถเกิดอะไรขึ้นได้ ซึ่งการเดินแต่ละครั้งอาจจะไม่เป็นผลดีต่อครั้งนั้นแต่อาจมีผลดีในครั้งต่อๆ ไปก็ได้

2.1.3 เทคนิค Classification [5]

เทคนิคการจำแนกประเภทข้อมูล เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ เพื่อแสดงให้เห็นความแตกต่างระหว่าง Class หรือ กลุ่มของข้อมูลได้ และเพื่อทำนายว่าข้อมูลนี้ ควรจัดอยู่ใน Class ใด ซึ่งโมเดลที่ใช้จำแนกข้อมูลออกเป็นกลุ่มตามที่ได้กำหนดไว้จะขึ้นอยู่กับการวิเคราะห์เซตของข้อมูลสอน (Training data) โดยนำ Training data มาสอนให้ระบบเรียนรู้ว่ามีข้อมูลใดอยู่ใน Class เดียวกันบ้าง โดยผลลัพธ์ที่ได้จากการเรียนรู้ คือ Classifier Model โดยที่โมเดลสามารถแทนได้ในหลายรูปแบบ เช่น Decision Tree และจะนำข้อมูลส่วนที่เหลือจาก Training data เป็นข้อมูลที่ใช้ทดสอบ (Testing data) ซึ่งเป็นกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบโดยข้อมูลนี้จะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลเพื่อทดสอบความถูกต้อง โดยสามารถปรับปรุงโมเดลจนกว่าจะได้ค่าความถูกต้องในระดับที่น่าพอใจ หลังจากนั้นเมื่อมีข้อมูลใหม่เข้ามา จะนำข้อมูลมาผ่านโมเดล โดยโมเดลจะสามารถทำนายกลุ่มของข้อมูลนี้ได้

2.1.3.1 Random Forest [6]

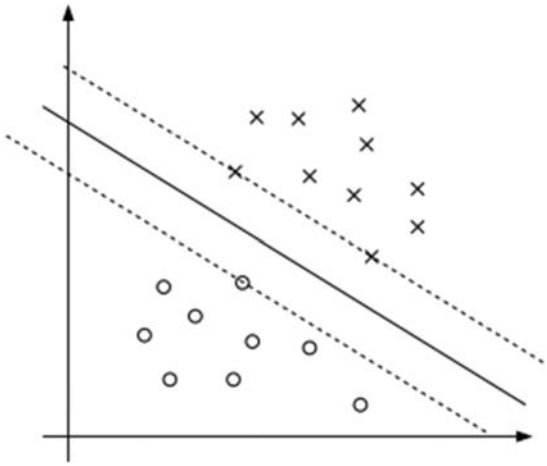
Random Forest เป็นโมเดลที่จะทำการสุ่มข้อมูล Training Data และสุ่มเลือกคุณลักษณะ (Attribute) ต่างๆ ออกมาหลายๆ ชุด มาทำการสร้างโมเดลด้วยเทคนิค Decision Trees หลายๆ ต้น แล้วนำโมเดลที่สร้างไปใช้ในการทำนายข้อมูล ซึ่งใน Decision Trees แต่ละต้นจะให้คำตอบออกมาในขั้นตอนสุดท้าย หลังจากนั้นจะเอาคำตอบของ Decision Trees แต่ละต้นมารวมกันเพื่อหาว่าคำตอบไหนได้ค่าที่เหมาะสมมากที่สุด โดยทั่วไป Random Forest จะให้ผลการทำนายที่ดีกว่าการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning)

2.1.3.2 Support Vector Machine [7]

Support Vector Machine หรือ SVM เป็นอัลกอริทึมในการคัดแยกกลุ่มเพื่อจัดประเภทหรือจำแนกประเภทข้อมูลที่มีการนำมาใช้ในด้านการประมวลผลภาพ เป็นวิธีการจำแนกกลุ่มข้อมูล ที่อาศัยระนาบการตัดสินใจที่เรียกว่า ระนาบเกิน หรือไฮเปอร์เพลน (Hyperplane) มาใช้ในการจำแนกกลุ่มข้อมูลโดยใช้สมการเส้นตรงในการแบ่งข้อมูลออกเป็น

2กลุ่มแยกออกจากกัน SVM มีรูปแบบในการเรียนรู้เป็นกระบวนการเลือกแบบจำลองที่เหมาะสมที่สุดจะทำให้ได้ค่าที่เหมาะสมที่สุดเป็นคำตอบ ดังนั้น SVM จึงเป็นที่นิยมและเริ่มนำไปใช้ในงานด้านการรู้จำรูปแบบซึ่งจะเลือกใช้ SVM แบบแบ่งกลุ่ม

Support Vector Machine สำหรับการแบ่งกลุ่มข้อมูลนั้นจะใช้ระนาบเกินที่เหมาะสมที่สุด (Optimal Hyperplane) ในการแบ่งกลุ่ม ในการสร้างระนาบเกินที่ใช้ในการแบ่งกลุ่มข้อมูลสามารถสร้างได้หลายแบบ แต่จะมีระนาบเกินที่เหมาะสมที่สุดเพียงระนาบเดียวเท่านั้น ที่สามารถรักษาระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่มที่ใกล้กันมากที่สุดได้



ภาพที่ 1 ตัวอย่าง Support Vector Classification
(ที่มา: https://www.jeremyjordan.me/content/images/2017/06/Srceen-Shot-2017-06-20-at-2.18.24-PM.png)

จากภาพที่ 1 ข้างต้น สามารถอธิบายได้ว่าจากการกระจายตัวของข้อมูลในรูป เราสามารถแบ่งแยกออกเป็น 2 กลุ่มได้อย่างชัดเจน ซึ่งปกติแล้วเราจะใช้สมการเส้นตรง (Liner model) เพื่อทำการแบ่งข้อมูลออกเป็น 2 กลุ่ม โดยเราจะสร้างเส้นแบ่งกลุ่ม (Hyperplane) ที่สามารถแบ่งกลุ่มด้วยระยะ Margin ที่มากที่สุด

2.1.3.3 Nearest Neighbors [8]

K-Nearest Neighbor (KNN) เป็นวิธีการจำแนกประเภทข้อมูล (Data Classification) วิธีการหนึ่ง โดยจัดเป็นวิธีการจำแนกประเภทข้อมูลแบบมีผู้ฝึกสอน (Supervised Learning) หรือการที่ทราบคำตอบของข้อมูลอยู่ก่อนแล้ว จากนั้น ใช้โมเดลในการจำแนกประเภท ข้อมูลจากข้อมูลฝึกที่ทราบคำตอบ วิธีการจำแนกจะใช้วิธีการวิเคราะห์จาก ข้อมูลที่

ใกล้เคียงที่สุดจำนวน K ตัว กับ ข้อมูลที่ต้องการจำแนกประเภทของข้อมูลหรือต้องการ ทำนายคลาสของข้อมูลใหม่โดยจะทำนายตามคลาสส่วนใหญ่ของข้อมูลฝึก K ตัว

2.1.4 เทคนิคการวิเคราะห์กลุ่ม (Clustering analysis) [9]

การวิเคราะห์กลุ่มซึ่งมีเป้าหมายเพื่อการจำแนกกลุ่มข้อมูลที่มีคุณลักษณะคล้ายกันอยู่ในกลุ่มเดียวกันโดยข้อมูลแต่ละกลุ่มจะถูกเรียกว่าคลัสเตอร์ (Cluster) การวิเคราะห์หรือจำแนกกลุ่มข้อมูลนั้น สามารถแบ่งออกได้เป็น 2 ประเภทได้แก่วิธีการแบบลำดับชั้น (Hierarchical algorithms) และวิธีการแบบไม่เป็นลำดับชั้น (Non-hierarchical algorithms)

2.1.4.1 Density-Based Methods [10]

เทคนิคการจัดกลุ่มข้อมูลโดยใช้วิธีการแบ่งข้อมูลเป็นส่วนๆ ส่วนใหญ่มักใช้การหาความสัมพันธ์ของแต่ละอ็อบเจ็กต์โดยการพิจารณาระยะทาง (Distance) ของอ็อบเจ็กต์เหล่านั้น ทำให้เทคนิคดังกล่าวมีข้อจำกัดคือสามารถค้นหาได้เฉพาะคลัสเตอร์ที่มีรูปทรงกลมเท่านั้นด้วย ข้อจำกัดดังกล่าวจึงนำไปสู่การพัฒนาเทคนิคใหม่ซึ่งพิจารณาความหนาแน่นของข้อมูลเป็นเกณฑ์ในการค้นหาคลัสเตอร์ หลักการทั่วไปของเทคนิคนี้คือการแผ่ขยายขอบเขตของคลัสเตอร์ไปเรื่อยๆ ตราบใดที่ความหนาแน่นของอ็อบเจ็กต์ มีค่าถึงค่าที่กำหนดนั่นคือแต่ละอ็อบเจ็กต์ของคลัสเตอร์ใดๆ จะต้องประกอบด้วยอ็อบเจ็กต์ซึ่งอยู่ใกล้กันภายในรัศมีที่กำหนด (Neighborhood) จำนวนไม่ น้อยกว่าค่าที่กำหนด ด้วยเทคนิคนี้สามารถใช้ในการกรองข้อมูลรบกวนซึ่งเป็นข้อมูลที่มีความหนาแน่นเบาบางได้ และยังสามารถค้นหาคลัสเตอร์ที่มีรูปทรงที่ซับซ้อนได้อีกด้วย

- **Density-based spatial clustering of applications with noise (DBSCAN) [11]** เป็นการหาบริเวณที่ข้อมูลเกาะกลุ่มกัน ซึ่งสามารถคำนวณได้จากข้อมูลที่อยู่รอบๆ ในรัศมีที่กำหนด การที่จะใช้ DBSCAN ได้ จำเป็นต้องมี 2 parameter ได้แก่
    1. Epsilon (eps) คือ รัศมีจากจุดศูนย์กลาง
    2. Min Points คือ จำนวนข้อมูลขั้นต่ำสำหรับการกำหนดศูนย์กลาง

ในแต่ละข้อมูลจะคำนวณหาข้อมูลที่ใกล้เคียงทั้งหมดในรัศมี eps ถ้าข้อมูล ไหนมีข้อมูลที่ใกล้เคียงมากกว่าหรือเท่ากับ Min Points ให้ข้อมูลนั้นเป็นข้อมูลหลัก และสร้างเป็นกลุ่มใหม่ และถ้าในแต่ละข้อมูลหลักมีข้อมูลที่ใกล้เคียงที่เชื่อมต่อกับอีก ข้อมูลหลักได้ให้รวมเป็นกลุ่มเดียวกันแต่ถ้าไม่เชื่อมต่อกันก็จะให้ข้อมูลนั้นเป็น Noise ซึ่งจะไม่อยู่ในกลุ่มใดๆเลย

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 งานวิจัยเรื่อง “Trajectory Clustering via Deep Representation Learning” [12]

Di Yao และคณะ ได้เสนอframeworkใหม่สำหรับการจัดกลุ่มเส้นโคจร โดยใช้อัลกอลิทึมการแยกพฤติกรรมการเคลื่อนที่เพื่อแยกพฤติกรรมที่เก็บลักษณะที่คงที่ของระยะทางและเวลาของเส้นทาง จากนั้นใช้ sequence to sequence auto-encoder เพื่อสร้างการแสดงผลของลำดับพฤติกรรมการเคลื่อนย้ายและระบุการคลาดเคลื่อนของระยะทางและเวลา ผู้วิจัยได้แสดงประสิทธิภาพของframeworkกับข้อมูลสังเคราะห์และข้อมูลจริง ผลการทดลองของเขาแสดงให้เห็นว่าวิธีการของผู้วิจัยมีความแม่นยำสูงกว่าวิธีการจัดกลุ่มอื่นๆบนข้อมูลสังเคราะห์ นอกจากนี้วิธีนี้สามารถนำไปใช้จัดกลุ่มเส้นโคจรและตรวจจับกลุ่มของวัตถุได้อย่างแม่นยำสำหรับข้อมูลจริง

2.2.2 งานวิจัยเรื่อง “Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web” [13]

Yu Zheng และคณะสนใจข้อมูล GPS มาทำการเรียนรู้เพื่อจำแนกประเภทการเดินทางแบบอัตโนมัติ ประเภทการเดินทางเช่น การเดิน การขับรถ เป็นต้น เพื่อแสดงนัยของของข้อมูลสามารถทำให้เข้าใจผู้ใช้ได้ และยังมีการใช้การคำนวณ context-aware จากประเภทการเดินทางปัจจุบันของผู้ใช้ และออกแบบเว็บไซต์

วิธีการของเขาประกอบด้วย 3 ส่วน คือ ส่วนแรกการแบ่งส่วนตามจุดเปลี่ยนและนำไปเปรียบเทียบกับโมเดลที่มีระยะเวลาสม่ำเสมอและความยาวสม่ำเสมอ ส่วนที่สองเป็นการใช้โมเดลที่แตกต่างกัน Decision Tree, Bayesian Net, Support Vector Machine (SVM) และ Conditional Random Field (CRF) ส่วนที่สาม คือ อัลกอริทึมที่เกิดหลังการทำงานที่ใช้ความน่าจะเป็นตามเงื่อนไขการประเมินผลใช้ข้อมูล

GPS ที่รวบรวมโดย 45 คนในช่วงเวลาหกเดือน ผลลัพธ์ที่ได้ การแบ่งส่วนตามจุดเปลี่ยนมีประสิทธิภาพในการทำนายสูงที่สุดและ Decision Tree มีประสิทธิภาพเหนือกว่าโมเดลอื่นๆ

2.2.3 งานวิจัยเรื่อง “Smartphone Inertial Sensor-Based Indoor Localization and Tracking With iBeacon Corrections” [14]

Zhenghua Chen และคณะ ได้เสนอการระบุตำแหน่งภายในอาคารที่ใช้เซนเซอร์จากโทรศัพท์มือถือกับ iBeacon โดยใช้วิธีการ Pedestrian Dead Reckoning (PDR) และ Kalman filter เข้ามาช่วย ซึ่งเมื่อนำไปทดลองเก็บข้อมูล และนำเปรียบเทียบกับการระบุตำแหน่งด้วย Wi-Fi fingerprint แล้ว วิธีการระบุตำแหน่งที่ผู้วิจัยนำเสนอให้ผลลัพธ์ที่ดีกว่า

2.2.4 งานวิจัยเรื่อง “Mining Individual Life Pattern Based on Location History” [15]

Yang Ye และคณะสนใจทำเหมืองหารูปแบบการใช้ชีวิตของแต่ละคนจากข้อมูลการระบุตำแหน่ง เขานำเสนอเค้าโครง LP-Mine เพื่อใช้ในการดึงข้อมูลรูปแบบการใช้ชีวิตอย่างมีประสิทธิภาพจากข้อมูลGPS ของแต่ละคน โดยเขานิยามให้รูปแบบการใช้ชีวิตขึ้นอยู่กับสถานที่สำคัญในการใช้ชีวิตของแต่ละคนและพิจารณาหลากหลายคุณสมบัติเพื่อจะรวมสถานที่สำคัญ LP-Mineแบ่งออกเป็นสองส่วนคือระยะ modeling และระยะ mining ในส่วนของระยะ modeling จะทำการเตรียมข้อมูลก่อนการดำเนินงานให้สามารถใช้เป็นอินพุตในระยะ mining ได้ และในส่วนของระยะ mining จะประยุกต์ใช้กลยุทธ์การแบ่งแยก เพื่อค้นหาความแตกต่างของแบบแผนแต่ละชนิด และนำมาทำการทดลองโดยการเก็บข้อมูลโดยอาสาสมัคร เพื่อยืนยันประสิทธิภาพของเค้าโครง

2.2.5 งานวิจัยเรื่อง “Beacon Placement for Range-Based Indoor Localization” [16]

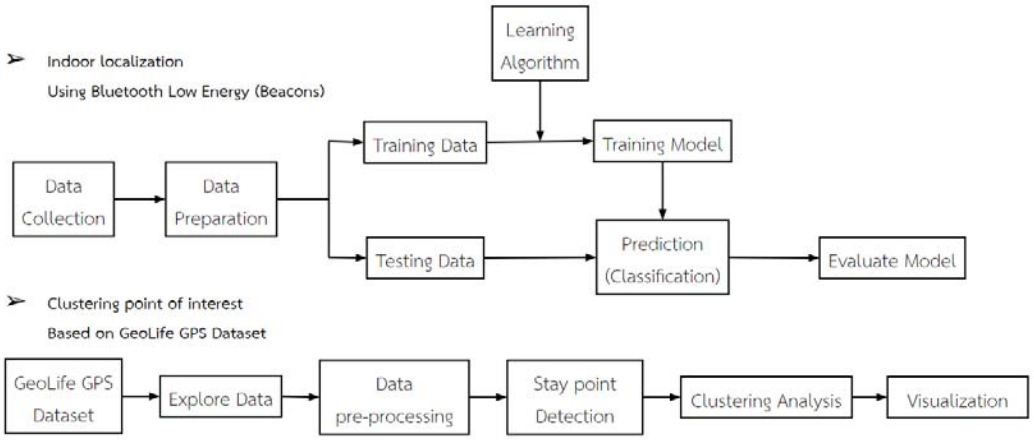
Niranjini Rajagopal และคณะ กล่าวถึงการระบุตำแหน่งที่อยู่ซึ่งในปัจจุบันเป็นสิ่งสำคัญสิ่งหนึ่งในชีวิตประจำวัน การระบุตำแหน่งที่อยู่โดยทั่วไปนิยมใช้ GPS ในการระบุตำแหน่ง แต่ GPS ไม่สามารถระบุตำแหน่งที่อยู่ภายในอาคารซึ่งมีสิ่งกีดขวางมากมายได้ สำหรับตำแหน่งในพื้นที่ภายในอาคารจำเป็นต้องใช้บีคอนส์เพื่อช่วยระบุตำแหน่งที่แม่นย่า ซึ่งการที่จะระบุตำแหน่งที่แม่นยำภายในอาคารได้จำเป็นต้องใช้บี

คอนส์อย่างน้อย 3 ตัว ส่งผลให้ค่าใช้จ่ายค่อนข้างสูง คณะผู้จัดทำวิจัยนี้จึงมีความมุ่งมั่นที่จะเพิ่มศักยภาพของการคำนวณและระบุตำแหน่งภายในอาคารที่ถูกต้องและแม่นย่าด้วยการลดจำนวนบีคอนส์ให้น้อยลงเพื่อลดค่าใช้จ่าย แต่ประสิทธิภาพในการระบุตำแหน่งยังคงเดิม

บทที่ 3
วิธีการดำเนินงาน

3.1 แผนการดำเนินงาน

โครงงานของคณะผู้จัดทำได้แบ่งออกเป็น 2 ส่วน โดยส่วนที่หนึ่งเป็นการระบุตำแหน่งภายในอาคาร และส่วนที่สองเป็นการค้นหาตำแหน่งที่ได้รับความสนใจ ดังที่แสดงในภาพที่ 2



ภาพที่ 2 แผนการดำเนินงานโดยรวม

3.1.1 การระบุตำแหน่งภายในอาคาร

- **ศึกษาค้นคว้าข้อมูลแนวคิดและทฤษฎีที่เกี่ยวข้องกับโครงงาน**

  เป็นขั้นตอนในการกำหนดหัวข้อโครงงานที่ทางผู้จัดทำสนใจ จากนั้นทำการศึกษางานวิจัย หลักการ ทฤษฎี เตรียมอุปกรณ์ เครื่องมือ โปรแกรม ที่ใช้ในการดำเนินโครงงาน

- **ศึกษาและเก็บค่าสัญญาณ RSSI จากอุปกรณ์เคลื่อนที่**

  ทำการศึกษาและทดลองเก็บค่าสัญญาณ RSSI ของอุปกรณ์เคลื่อนที่ โดยใช้อุปกรณ์เคลื่อนที่รับความเข้มของสัญญาณจากตัวปล่อยสัญญาณบีคอนส์ จากนั้นนำข้อมูลที่ได้อัพขึ้น Firebase แบบ Real-time

- **เตรียมข้อมูล**

  แบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 80% ต่อ 20% ซึ่งข้อมูลส่วนที่หนึ่ง หรือ Training Data (80%) ใช้ในการสร้างโมเดล และข้อมูลส่วนที่สอง หรือ Testing Data (20%) ใช้ในการทดสอบประสิทธิภาพของโมเดล

- **การพัฒนาแบบจำลอง (Modeling) ด้วยอัลกอริทึมต่างๆ**

  สร้าง Model ด้วย Classification Algorithm ทั้ง 3 แบบ ได้แก่ K-Nearest Neighbor, Support Vector Classification และ Random Forest พร้อมทั้งทำการประเมินความถูกต้องด้วยการหาค่า Accuracy และค่าความคลาดเคลื่อนเฉลี่ย

- **สรุปและเผยแพร่งานวิจัย**

3.1.2 การค้นหาตำแหน่งที่ได้รับความสนใจ

- **ค้นหาชุดข้อมูลและศึกษาโค้ดที่เกี่ยวข้องกับ Machine Learning ที่ใช้กับการระบุตำแหน่งที่ได้รับความสนใจ**

  เป็นการค้นหาและศึกษาชุดข้อมูลที่เหมาะสมกับการทำ Machine Learning กับเรื่องที่ผู้จัดทำสนใจ

- **สำรวจข้อมูลและเตรียมข้อมูล**

  นำชุดข้อมูล Geolife GPS trajectory ที่สืบค้นได้ มาทำการสำรวจรูปแบบข้อมูล และทำการเตรียมข้อมูลโดยคำนวณความเร็วของระหว่างตำแหน่ง และลบข้อมูลที่มีความเร็วเกิน 100 เมตรต่อวินาที

- **ศึกษาและวิเคราะห์ข้อมูลที่ได้จากอัลกอริทึมต่างๆ**

  หาตำแหน่งที่ได้รับความสนใจของแต่ละผู้ใช้โดยนำข้อมูลที่ทำการเตรียมข้อมูลแล้ว มาทำการหาตำแหน่งหยุดด้วยอัลกอริทึม หลังจากนั้นนำตำแหน่งหยุดที่มาทำการ DBSCAN เพื่อจัดกลุ่มตำแหน่งหยุดและหาจุดกึ่งกลางของกลุ่มซึ่งจะเป็นตำแหน่งที่ได้รับความสนใจ นำมาเปรียบเทียบหาระยะทางกับสถานที่ 12 สถานที่แล้วเลือกสถานที่ที่ระยะทางใกล้ที่สุด

- **ประเมินผล**

  นำข้อมูลที่วิเคราะห์เสร็จแล้วมาประเมินผล โดยดูระยะทางที่ใกล้ที่สุด น้อยกว่าหรือเท่ากับ 2 กิโลเมตร ให้เป็นการหาตำแหน่งที่ได้รับความสนใจที่ถูกต้อง และระยะทางที่ใกล้ที่สุดมากกว่า 2 กิโลเมตร เป็นตำแหน่งที่ได้รับความสนใจที่ไม่ถูกต้อง

- **สรุปและเผยแพร่งานวิจัย**

3.2 แผนการดำเนินงานตลอดโครงงาน

ขั้นตอนการดำเนินการ	2561				2562				
	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
เริ่มต้นและวางแผนโครงการ									
วางแผนจัดทำโครงการ									
หาข้อมูลศึกษาแนวคิดและทฤษฎี									
ศึกษางานวิจัยที่เกี่ยวข้อง									
ศึกษาการเก็บค่าสัญญาณจากบีคอนส์									
ศึกษาและทำการทดลองจัดเก็บค่าสัญญาณ RSSI ของอุปกรณ์ที่มีสัญญาณ Bluetooth จากบีคอนส์									
ศึกษาเกี่ยวกับ Machine Learning									
ศึกษาข้อมูลเกี่ยวกับ Classification									
ศึกษาข้อมูลเกี่ยวกับการระบุตำแหน่งที่ได้รับความสนใจ									
การพัฒนาแบบจำลอง (Modeling)									
การระบุตำแหน่งที่									

ได้รับความสนใจ									
หาข้อผิดพลาดและแก้ไขปรับปรุง									
สรุปและเผยแพร่งานวิจัย									

3.3 อุปกรณ์และเครื่องมือที่ใช้

3.3.1 ฮาร์ดแวร์ (Hardware)

- **คอมพิวเตอร์ PC**
  - Intel® CORE™ i7 CPU M370 @ 2.40GHz
  - RAM 8 GB
  - Windows 10 64-bit
- **สมาร์ทโฟนแอนดรอยด์**
  - Samsung galaxy note 3
  - Android 5.0 (Lollipop)
  - RAM 3 GB
  - A15 1.9 GHz + A7 1.3 GHz Octa Core
- **Estimote Beacon**
  - Bluetooth 4.0 LE : ใช้สื่อสารข้อมูลระหว่างตัวบีคอนส์กับแอปพลิเคชันบนมือถือ
  - ARM Cortex M0 Processor : CPU ใช้ควบคุมการรับส่งข้อมูลของอุปกรณ์ รวมถึง การปรับแต่งค่าของตัวบีคอนส์ เช่น ระยะเวลาในการส่งสัญญาณ
  - Coin Battery : ถ่านแบบเม็ดกระดุม

3.3.2 ซอร์ฟแวร์ (Software)

- Google Collaboratory
- Android studioเป็น IDE Tool จาก Google สำหรับพัฒนา Android โดยเฉพาะ

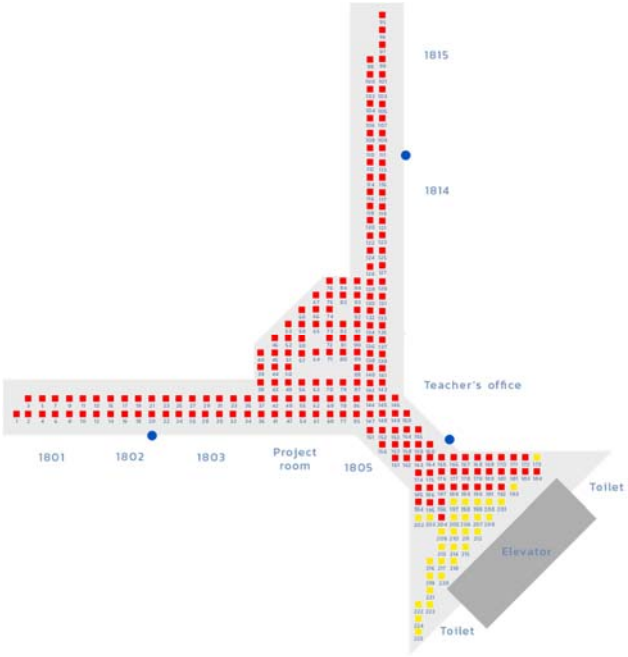
3.3.3 ภาษาที่ใช้ (Language)

- Python
- Java

บทที่ 4
ผลการดำเนินงาน

4.1 ผลลัพธ์จากการเก็บข้อมูล

การเก็บข้อมูลด้วยวิธีการรับสัญญาณจากบีคอนส์ซึ่งเราได้ทำการเก็บข้อมูลบริเวณชั้น 18 ของอาคารคณะวิทยาศาสตร์ โดยจะเก็บข้อมูลทั้งหมด 225 จุด เนื่องจากบีคอนส์มีจำนวนน้อยและสัญญาณไม่เพียงพอจึงเก็บข้อมูลได้เพียง 195 จุด ตามที่แสดงในภาพที่ 3 โดยจุดสีแดงแสดงตำแหน่งที่สามารถเก็บข้อมูลได้และจุดสีเหลืองแสดงตำแหน่งที่ไม่สามารถเก็บข้อมูลได้ ซึ่งข้อมูลที่เก็บได้แสดงดังภาพที่ 4



ภาพที่ 3 แผนผังที่ทำการเก็บข้อมูลบริเวณชั้น 18

timestamp	rssi1	rssi2	rssi3	x	y	block
2019-01-28 13:36:01.362	-83	-83	-95	1.05	16.80	1
2019-01-28 13:36:02.479	-83	-81	-93	1.05	16.80	1
2019-01-28 13:36:05.848	-77	-86	-96	1.05	16.80	1
2019-01-28 13:36:06.962	-83	-82	-89	1.05	16.80	1
2019-01-28 13:36:08.092	-80	-89	-90	1.05	16.80	1
2019-01-28 13:36:09.218	-78	-84	-91	1.05	16.80	1
2019-01-28 13:36:10.345	-76	-85	-91	1.05	16.80	1
2019-01-28 13:36:11.459	-79	-85	-91	1.05	16.80	1
2019-01-28 13:36:14.845	-83	-87	-94	1.05	16.80	1
2019-01-28 13:36:24.919	-75	-88	-90	1.05	16.80	1
2019-01-28 13:36:26.030	-74	-89	-88	1.05	16.80	1
2019-01-28 13:36:27.143	-82	-95	-100	1.05	16.80	1
2019-01-28 13:36:29.385	-85	-94	-93	1.05	16.80	1
2019-01-28 13:36:30.515	-84	-94	-94	1.05	16.80	1

ภาพที่ 4 ตัวอย่างข้อมูลที่เก็บได้

โดย timestamp คือ วันและเวลาที่เก็บข้อมูล

x และ y คือ ตำแหน่งของจุดที่วัดค่าสัญญาณ

rssi1, rssi2, rssi3 คือ ค่าความแรงของสัญญาณบีคอนส์ตัวที่ 1, 2, 3

Block คือ เลขประจำตำแหน่งของจุดที่วัดค่าสัญญาณ

ซึ่งค่าความแรงของสัญญาณที่ได้รับมาทั้งหมด 3,212 จุด ในแต่ละจุดจะมีความแตกต่างกันเล็กน้อย และอาจมีบางครั้งที่ค่าแปรปรวน ต่อมาเราได้นำข้อมูลทั้งหมด มาวิเคราะห์กับอัลกอริทึม K-Nearest Neighbor , Random Forest และ Support Vector Classification โดยแบ่งข้อมูลเป็นTrain data 80% และ Test data 20%

4.2 ผลลัพธ์จากการคำนวณข้อมูล

ทำการทำนายตำแหน่งได้ตัวอย่างผลลัพธ์ออกมาแสดงดังภาพที่ 5

Actual	KNN	RF	SVC
31	31	47	49
87	81	81	87
13	13	13	13
126	43	125	125
178	166	166	166
132	66	66	53
72	132	137	72
70	56	64	56
112	111	107	111
39	49	49	49

ภาพที่ 5 ผลการคำนวณหาตำแหน่ง

โดยที่ Actual คือ ตำแหน่งที่อยู่จริงขณะที่ทำการทดลอง และ KNN, RF และ SVC คือ ตำแหน่งที่อัลกอริทึมแต่ละแบบคำนวณได้

ทำการเก็บข้อมูลแบบเส้นทางเดินตั้งแต่บริเวณหน้าห้อง 1802 ถึงบริเวณทางเข้าห้องพักอาจารย์ และนำมาทำนายด้วยอัลกอริทึมแต่ละแบบได้ผลลัพธ์ภาพที่ 6

a) KNN

b) RF

c) SVC

ภาพที่ 6 ผลลัพธ์การทำนายข้อมูลแบบเส้นทางเดินโดยใช้อัลกอริทึมแต่ละแบบ

โดย เส้นสีน้ำเงิน คือ เส้นทางการเดินจริง เส้นสีแดง คือ เส้นทางเดินที่ทำนายโดยอัลกอริทึมแต่ละแบบ และเส้นสีเขียว คือ เส้นทางเดินที่ทำนายได้ซึ่งนำไปผ่าน Kalman filter

4.3 ผลลัพธ์ที่ได้จากการทดสอบ

จากการทดลองกับอัลกอริทึมทั้ง 3 แบบ ได้ผลการทดสอบออกมาดังตารางที่ 1 โดยจากการทำนายด้วยอัลกอริทึมทั้ง 3 แบบ จะได้ตำแหน่งและนำตำแหน่งที่ได้มาคำนวณหาความคลาดเคลื่อนโดยใช้สูตรดังนี้

$$\frac{\Sigma\sqrt{(x_{actual}-x_{predict})^2-(y_{actual}-y_{predict})^2}}{N}$$

โดย x_{actual} คือค่าในแนวแกน x จริง , $x_{predict}$ คือค่าในแนวแกน x ที่ทำนายได้

y_{actual} คือค่าในแนวแกน y จริง , $y_{predict}$ คือค่าในแนวแกน y ที่ทำนายได้

N คือจำนวนตำแหน่งทั้งหมด

เทคนิคการแบ่งกลุ่ม	ความแม่นยำในการระบุตำแหน่ง	ความคลาดเคลื่อนเฉลี่ยในการระบุตำแหน่ง (เมตร)
K-Nearest Neighbor	19.91%	5.04
Support Vector Classification	21.62%	4.45
Random Forest	19.28%	4.70

ตารางที่ 1 ผลการทดสอบความแม่นยำและความคลาดเคลื่อน

จากตารางค่าความแม่นยำในการแบ่งกลุ่มของวิธี Support Vector Classification มีความแม่นยำสูงสุด คือ 21.62% แม้ว่าจะมีความแม่นยำมากที่สุดจากวิธีการทั้งหมดแต่ก็ยังถือว่าน้อยมากเนื่องจากข้อมูลที่ทำการเก็บรวบรวมมีจำนวนน้อยและวิธี Support Vector Classification ก็ยังมีความคลาดเคลื่อนเฉลี่ยในการระบุตำแหน่งน้อยที่สุด คือ 4.45 เมตร ซึ่งความคลาดเคลื่อนเฉลี่ยน้อยมีผลมาจากตำแหน่งของจุดที่เก็บข้อมูลใกล้กันเกินไป

4.4 ผลลัพธ์สำรวจและทำการเตรียมข้อมูล

จากการสำรวจชุดข้อมูล Geolife GPS trajectory ได้ผลลัพธ์ดังตารางที่ 2 และตัวอย่างข้อมูลดังภาพที่ 7

ชุดข้อมูลที่ใช้	
จำนวนผู้ใช้	182

จำนวนเส้นทาง	18,670
จำนวนตำแหน่ง	24,876,978

ตารางที่ 2 แสดงรายละเอียดชุดข้อมูล Geolife GPS trajectory

latitude	longitude	zero	altitude	number of days	Date	Time
39.984702	116.318417	0	492	39744.120185	2008-10-23	02:53:04
39.984683	116.318450	0	492	39744.120255	2008-10-23	02:53:10
39.984686	116.318417	0	492	39744.120313	2008-10-23	02:53:15
39.984688	116.318385	0	492	39744.120370	2008-10-23	02:53:20
39.984655	116.318263	0	492	39744.120428	2008-10-23	02:53:25
39.984611	116.318026	0	493	39744.120486	2008-10-23	02:53:30

ภาพที่ 7 ตัวอย่างข้อมูลเส้นทางเดินของผู้ใช้

โดย Latitude คือ พิกัดที่ใช้บอกตำแหน่งบนพื้นโลก โดยวัดจากเส้นศูนย์สูตร

Longitude คือ พิกัดที่ใช้บอกตำแหน่งบนพื้นโลก โดยวัดไปทางตะวันออกหรือตะวันตก

Zero คือ ข้อมูลทั้งหมดถูกตั้งค่าเป็น 0

Altitude คือ ระดับความสูงจากระดับน้ำทะเลขึ้นไปในแนวดิ่ง หน่วยเป็น ฟุต

Number of days คือ จำนวนวันที่ผ่านมาตั้งแต่วันที่ 30/12/1899 (Fractional)

Date คือ วัน และ Time คือ เวลา

ในการดำเนินงานเลือกใช้ GPS trajectories ของผู้ใช้ 10 คน จากชุดข้อมูล Geolife GPS trajectory เพื่อทำการทดสอบ และคำนวณความเร็วระหว่างตำแหน่ง ลบข้อมูลที่มีความเร็วมากกว่า 100 เมตร/วินาที เนื่องจากความเร็วในการเคลื่อนที่ไม่มีโอกาสเกิน 100 เมตร/วินาที โดยมีรายละเอียดดังตาราง

ชุดข้อมูลที่ใช้	
จำนวนผู้ใช้	10
จำนวนเส้นทาง	1,475
จำนวนตำแหน่ง	2,672,905
จำนวนตำแหน่งหลังทำความสะอาดข้อมูล	2,662,007

ตารางที่ 3 แสดงรายละเอียดชุดข้อมูล Geolife GPS trajectory ที่นำมาใช้

ตารางที่ 3 แสดงจำนวนตำแหน่งก่อนและหลังทำความสะอาดข้อมูล ซึ่ง 10,898 ตำแหน่ง ถูกลบออกจากชุดข้อมูลเดิม

4.5 ผลลัพธ์การหาตำแหน่งหยุด (Stay point) ของแต่ละผู้ใช้

ตารางที่ 4 แสดงจำนวนตำแหน่งหยุดที่คำนวณได้จากผู้ใช้แต่ละคนหลังจากทำงานอัลกอริทึม โดยผลลัพธ์แสดงบนแผนที่ดังภาพที่ 8 ใช้ thresholds ในการหาตำแหน่งหยุด คือ 200 เมตร ใน 20 นาที

ผู้ใช้	จำนวนเส้นทาง	จำนวนตำแหน่ง	จำนวนตำแหน่งหลังทำความสะอาดข้อมูล	จำนวนตำแหน่งหยุด
1	171	173,870	173,868	424
2	175	248,217	248,217	312
3	322	485,226	484,084	995
4	395	439,397	439,397	1,220
5	86	109,046	109,046	176
6	28	31,830	31,830	36
7	54	87,217	87,217	139
8	34	77,910	77,910	78
9	49	84,616	84,616	135
10	161	935,576	925,822	101
รวม	**1,475**	**2,672,905**	**2,662,007**	**3,616**

ตารางที่ 4 แสดงรายละเอียดตำแหน่งหยุด

a) ผู้ใช้ที่ 1 b) ผู้ใช้ที่ 2

ภาพที่ 8 ตำแหน่งหยุด

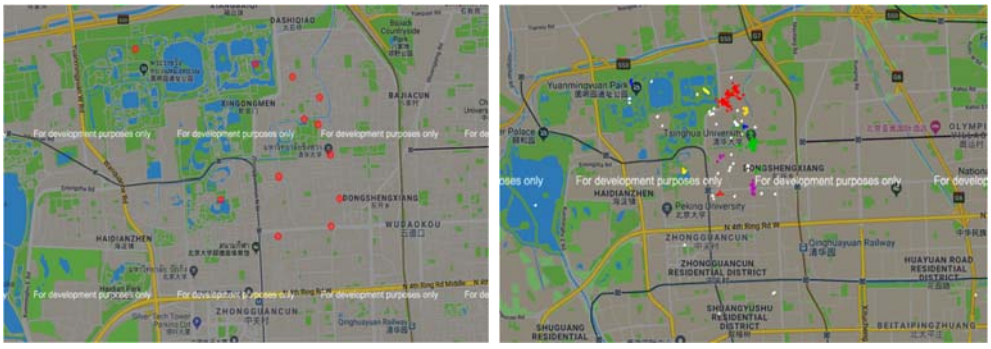
4.6 ผลลัพธ์การนำตำแหน่งหยุดมาจัดกลุ่มโดยใช้ DBSCAN

จากตำแหน่งหยุดที่ได้ในข้อ 4.5 นำมาทำอัลกอริทึม DBSCAN ของแต่ละผู้ใช้ โดยใช้ eps = 100 เมตร และ MinPoints = 4 ซึ่งเมื่อเปรียบเทียบโดยใช้ eps = 100 เมตร และ MinPoints = 4 มีค่าต่างกันเพียงเล็กน้อย เพื่อหาตำแหน่งที่ได้รับความสนใจ

ผู้ใช้	จำนวนตำแหน่งหยุด	จำนวนตำแหน่งหยุดที่ถูกเลือก	วนตำแหน่งที่ได้รับความส
1	424	323	13
2	312	240	9
3	995	798	22
4	1,220	1,016	21
5	176	149	8
6	36	14	1
7	139	78	3
8	78	35	3
9	135	126	4
10	101	9	2
รวม	**3,616**	**2,788**	**86**

ตารางที่ 5 แสดงจำนวนตำแหน่งที่ได้รับความสนใจของแต่ละผู้ใช้โดย DBSCAN
(Eps = 100 เมตร และ MinPoints = 4)

จากผลลัพธ์ที่แสดงในตารางที่ 5 การจัดกลุ่มตำแหน่งหยุดของแต่ละผู้ใช้ ดังภาพที่ 9



a) การจัดกลุ่ม b) ตำแหน่งที่ได้รับความสนใจ

ภาพที่ 9 ผลลัพธ์ของผู้ใช้ที่ 1

4.7 ผลลัพธ์การนำตำแหน่งที่ได้รับความสนใจมาประเมินความถูกต้อง

จากตำแหน่งที่ได้รับความสนใจนำมาประเมินความถูกต้อง โดยหาระยะทางจากจุดศูนย์กลางของแต่ละกลุ่มกับตำแหน่งที่ได้รับความสนใจของผู้ใช้ที่กำหนดไว้ ได้แก่ Tsinghua University, Yuanmingyuan Park, Beijing University, Microsoft Research Asia, Baiwanzhuang South Street No.10 Courtyard, Beijing Power Supply Training School, Tianyu Building, Zizhuyuan Park Amusement Park, Renmin University of China, Yangchun Guanghua Homeland Fengshuyuan, Beijing Institute of Technology Gymnasium, University of Chinese Academy of Sciences และ Beihang University โดยเลือกระยะทางที่ใกล้ที่สุด ถ้าน้อยกว่าหรือเท่ากับ 2 กิโลเมตร ให้เป็นการหาตำแหน่งที่ได้รับความสนใจที่ถูกต้อง และมากกว่า 2 กิโลเมตร เป็นตำแหน่งที่ได้รับความสนใจที่ไม่ถูกต้อง ซึ่งได้ผลลัพธ์ดังตารางที่ 6

ผู้ใช้	นวนตำแหน่งที่ได้รับความสน	จำนวนตำแหน่งที่ได้รับความสนใจที่ถูกต้อง
1	13	12
2	9	6
3	22	18
4	21	14
5	8	6
6	1	1
7	3	2
8	3	1
9	4	2
10	2	1
รวม	**86**	**63**

ตารางที่ 6 แสดงจำนวนตำแหน่งที่ได้รับความสนใจที่ถูกต้องของแต่ละผู้ใช้

จากผลลัพธ์ที่แสดงในตารางที่ 6 แสดงตำแหน่งที่ได้รับความสนใจที่ถูกต้อง ซึ่งจากตำแหน่งที่ได้รับความสนใจทั้งหมดคือ 86 ตำแหน่ง และสามารถหาตำแหน่งที่ถูกต้องได้ 46 ตำแหน่ง คิดเป็น 73.26%

บทที่ 5
สรุปผลการดำเนินงานและข้อเสนอแนะ

5.1 สรุปผลการดำเนินโครงงาน

จากผลลัพธ์การเก็บข้อมูลค่า RSSI เก็บข้อมูลได้ไม่ครบถ้วนตามที่กำหนดตำแหน่งไว้ เนื่องจากสัญญาณบีคอนส์มีความแรงไม่พอ และเมื่อนำมาทำการเตรียมข้อมูลแล้วนำไปผ่านการเรียนรู้ของเครื่อง และมาทำการทำนายตำแหน่ง ซึ่งระบุตำแหน่งภายในอาคารได้แต่ไม่แม่นยำ แต่ค่าเฉลี่ยความคลาดเคลื่อนในการระบุตำแหน่งมีค่าน้อยเหมาะสมกับชุดข้อมูลที่มีจำนวนน้อย และอัลกอริทึม Support Vector Classification ให้ความแม่นยำในการระบุตำแหน่งสูงสุด ในอนาคตผู้วิจัยคาดหวังที่จะเก็บข้อมูลเพิ่มเติม ซึ่งอาจส่งผลให้มีความแม่นยำสูงขึ้นและสามารถใช้ในการระบุตำแหน่งได้

จากผลลัพธ์ในส่วนของการค้นหาตำแหน่งที่ได้รับความสนใจแสดงให้เห็นว่าทำอย่างไรที่จะจัดการข้อมูลดิบของ GPS ได้ โดยใช้การทำเหมืองข้อมูลเพื่อค้นหาประโยชน์และสิ่งที่น่าสนใจจากเส้นทางและผู้ใช้ และตำแหน่งที่ได้รับความสนใจจะสอดคล้องกับสถานที่ที่ต่างกันขึ้นอยู่กับแหล่งที่มาของ GPS เนื่องจาก Geolife GPS trajectory เป็นชุดข้อมูลที่ส่วนใหญ่บันทึกโดยนักวิชาการ ตำแหน่งที่ได้รับความสนใจจึงเป็นมหาวิทยาลัย อาคาร ที่อยู่อาศัยหรือร้านอาหารในมหาวิทยาลัย รวมทั้งศูนย์วิจัย นอกจากนี้วิธีการนี้สามารถนำไปใช้กับแหล่งข้อมูล GPS อื่นๆ ได้ ตัวอย่างเช่นหากนำมาใช้สำหรับ GPS ของนักท่องเที่ยว ตำแหน่งที่ได้รับความสนใจสามารถสอดคล้องกับสถานที่ท่องเที่ยวโรงแรมหรือร้านอาหาร เป็นต้น

5.2 ปัญหาและอุปสรรคในการดำเนินโครงงาน

1. การเก็บข้อมูลแต่ละจุดในเวลานาน
2. ข้อมูลที่เก็บในแต่ละจุดมีจำนวนน้อยทำให้ความแม่นยำไม่สูงพอ
3. บริเวณหน้าลิฟต์ไม่สามารถรับสัญญาณจากบีคอนส์ได้ครบทั้งสามตัว จึงไม่สามารถเก็บข้อมูลได้
4. ชุดข้อมูลที่นำมาใช้ไม่ระบุชื่อสถานที่

5.3 ข้อเสนอแนะ

ในอนาคตหากมีระยะเวลาในการทำงานมากขึ้นผู้วิจัยจะสามารถพัฒนาระบบให้มีความแม่นยำมากขึ้นโดยการติดตั้งอุปกรณ์ส่งสัญญาณบีคอนส์เพิ่ม และทำการเก็บข้อมูลเพิ่มเป็นจำนวนมากทั้งแบบเฉพาะตำแหน่งและเส้นทางภายในอาคาร ส่งผลให้ระบบมีประสิทธิภาพมากยิ่งขึ้น และ ในส่วนของตำแหน่งที่ได้รับความสนใจ ทำการหาตำแหน่งที่ได้รับความสนใจของผู้ใช้มากกว่า 10 คน และนำตำแหน่งที่ได้รับความสนใจของผู้ใช้ทั้งหมดมาหาตำแหน่งที่ได้รับความสนใจสูงสุด 5 อันดับแรก

บรรณานุกรม

[1] Curran, Kevin; Furey, Eoghan; Lunney, Tom; Santos, Jose; Woods, Derek; McCaughey, Aiden (2011). "**An Evaluation of Indoor Location Determination Technologies**". Journal of Location Based Services. 5 (2): 61–78.

[2] Michels, D. (2017, July 19). **3 ways to track people using location-based services.** Retrieved November 24, 2018, from https://www.networkworld.com/article/3208767/internet-of-things/3-ways-to-track-people-using-location-based-services.html

[3] ภาคย์ สธนเสาวภาคย์และชัชชัย คุณบัว. (2552). **ระบบระบุตำแหน่งวัตถุภายในอาคารโดยใช้มาตรฐาน IEEE 802.15.4.** สืบค้นเมื่อ 18 พฤศจิกายน 2561. จาก http://www.ecti-thailand.org/assets/papers/4_pub_1.pdf

[4] **การเรียนรู้ของเครื่อง.** สืบค้นเมื่อ 16 มกราคม 2562. จาก https://th.wikipedia.org/wiki/การเรียนรู้ของเครื่อง

[5] Ryoma. (2556). **Data Classification คืออะไร.** สืบค้นเมื่อ 16 มกราคม 2562. จาก http://l3thebomz.blogspot.com/2013/07/data-classification.html

[6] เอกสิทธิ์ พัชรวงศ์ศักดา. **การสร้างโมเดล Ensemble แบบต่างๆ.** สืบค้นเมื่อ 16 มกราคม 2562. จาก http://dataminingtrend.com/2014/data-mining-techniques/ensemble-model/

[7] นางสาวรติพร จันทร์กลั่น. (2557). **การปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนสำหรับการจำแนกข้อมูลภาพไบโอเมตริกซ์.** สืบค้นเมื่อ 21 มกราคม 2562. จาก https://core.ac.uk/download/pdf/70943503.pdf

[8] นายพงศกร ธีรรัศมี. (2558). **วิธีการหาค่า เค ที่เหมาะสมในการจำแนกแบบเคเนียร์เรสเนเบอร์กับข้อมูลทางการแพทย์.** สืบค้นเมื่อ 23 มกราคม 2562. จาก http://sutir.sut.ac.th:8080/sutir/bitstream/123456789/6072/2/Fulltext.pdf

[9] วีระยุทธ พิมพาภรณ์และพยุง มีสัจ. (2557). **การเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยวิธีการเลือกลักษณะสำคัญแบบพลวัตเพื่อเพิ่มประสิทธิภาพของอลักอริทึมการจัดกลุ่มบนปริภูมิย่อย A Comparative Efficiency of Clustering Using Dynamic Feature Selection Optimization of Subspace Clustering Algorithms.** สืบค้นเมื่อ 29 มกราคม 2562, จาก http://www2.it.kmutnb.ac.th/journal/pdf/vol20/ch06.pdf

[10] นายธรรมศักดิ์ เธียรนิเวศน์. (2548). **การลดขนาดข้อมูลด้วยน้ำหนักความหนาแน่นเพื่อการจัดกลุ่มข้อมูลขนาดใหญ่.** สืบค้นเมื่อ 29 มกราคม 2562, จาก http://www2.it.kmutnb.ac.th/journal/pdf/vol20/ch06.pd

[11] Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu. (1996). **A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.** Retrieved November 23, 2018, from https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

[12] Yao, D., Zhang, C., Zhu, Z., Huang, J., & Bi, J. (2017). **Trajectory clustering via deep representation learning.** 2017 International Joint Conference on Neural Networks (IJCNN), 3880-3887.

[13] Yu Zheng-Like Liu-Longhao Wang-Xing Xie. (2008). **Learning transportation mode from raw gps data for geographic applications on the web.** Proceeding of the 17th international conference on World Wide Web.

[14] Zhenghua Chen-Qingchang Zhu-Yeng Soh. (2016). **Smartphone Inertial Sensor-Based Indoor Localization and Tracking With iBeacon Corrections.** IEEE Transactions on Industrial Informatics.

[15] Yang Ye-Yu Zheng-Yukun Chen-Jianhua Feng-Xing Xie. (2009). **Mining Individual Life Pattern Based on Location History.** 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware.

[16] Rajagopal, N., Chayapathy, S., Sinopoli, B., & Rowe, A. (2016). **Beacon placement for range-based indoor localization.** 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN).

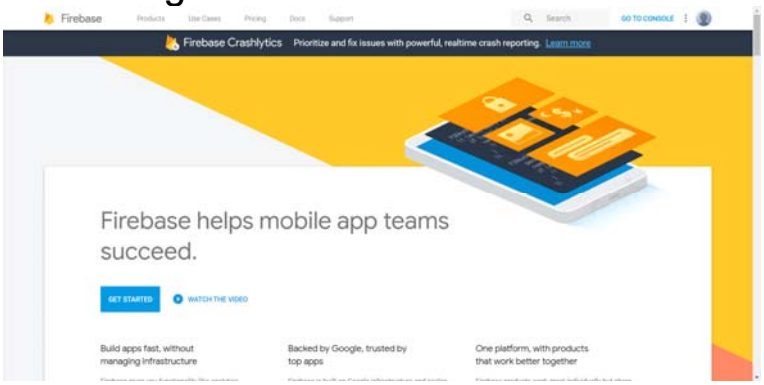
[17] Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., & Ma, W. (2008). **Mining user similarity based on location history**. Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS 08.

ภาคผนวก

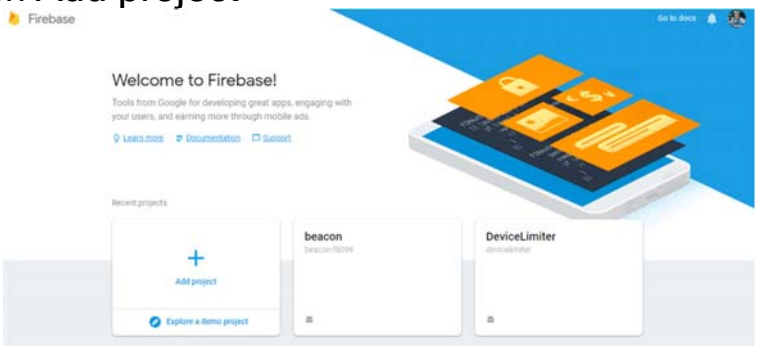
ภาคผนวก ก
สอนการติดตั้ง Firebase

การสมัครใช้ Firebase

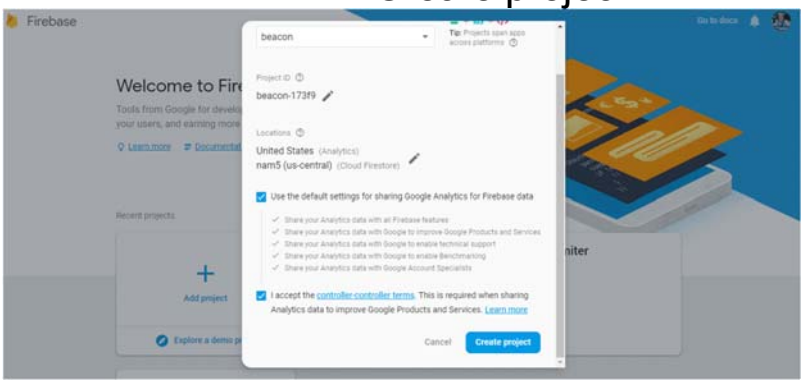
1. เข้าเว็บไซต์ https://firebase.google.com/ เข้าสู่ระบบด้วยอีเมลของ Google จากนั้นคลิกที่ GO TO CONSOLE



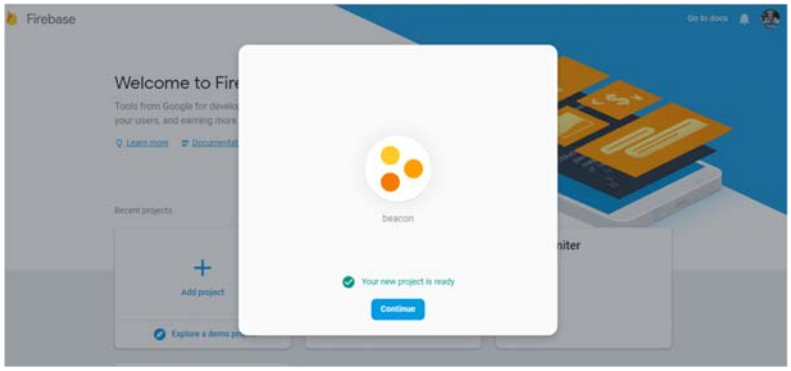
2. คลิกที่ Add project



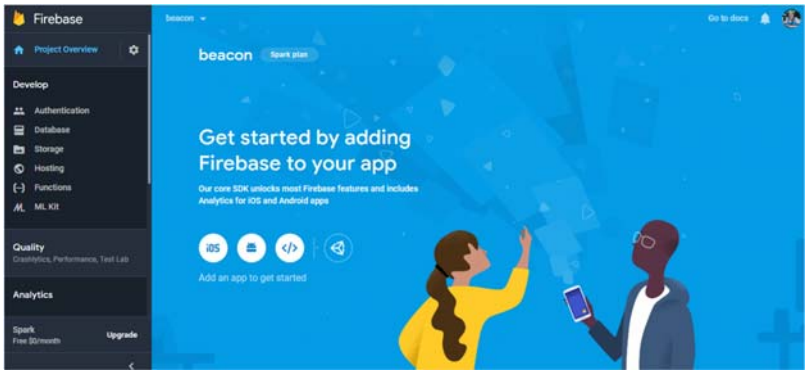
3. ตั้งชื่อโปรเจค จากนั้นคลิกที่ Create project



4. คลิกที่ Continue

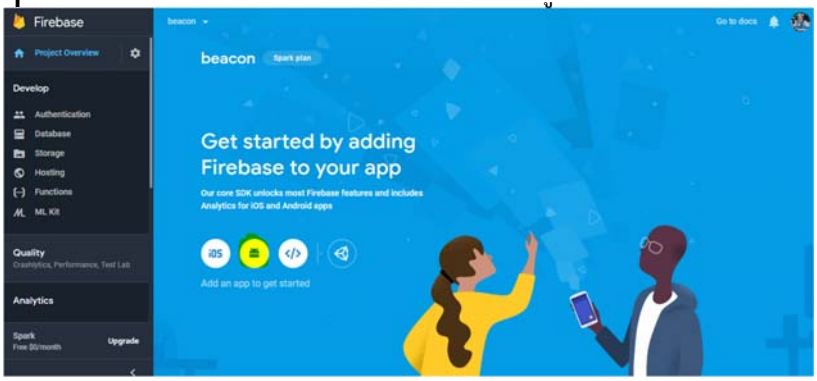


5. เสร็จสิ้นการติดตั้งโปรแกรม

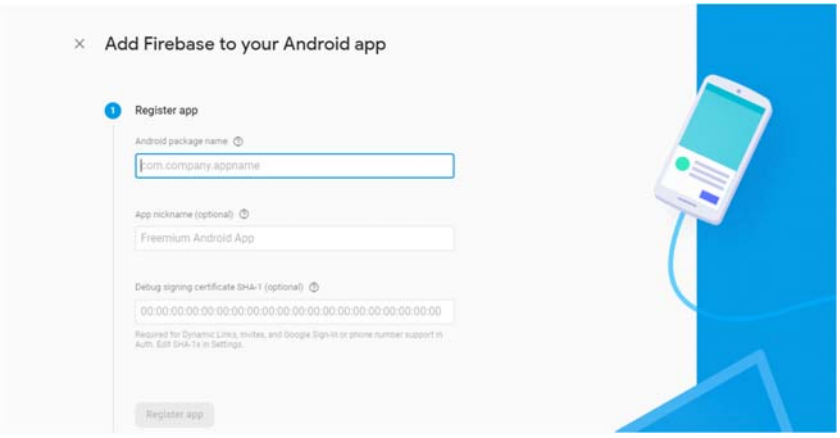


Set Up a Firebase in Android Project

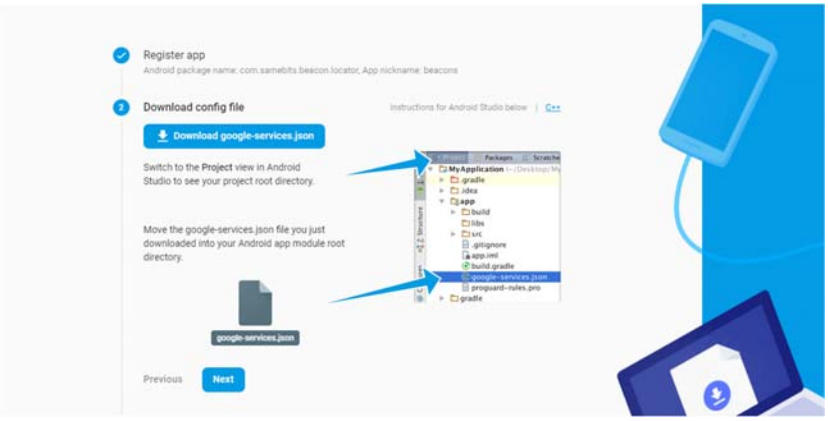
1. หลังจากสร้างโปรเจค Firebase เสร็จ ให้ Add Android application โดยกดบริเวณสีเหลืองดังรูป



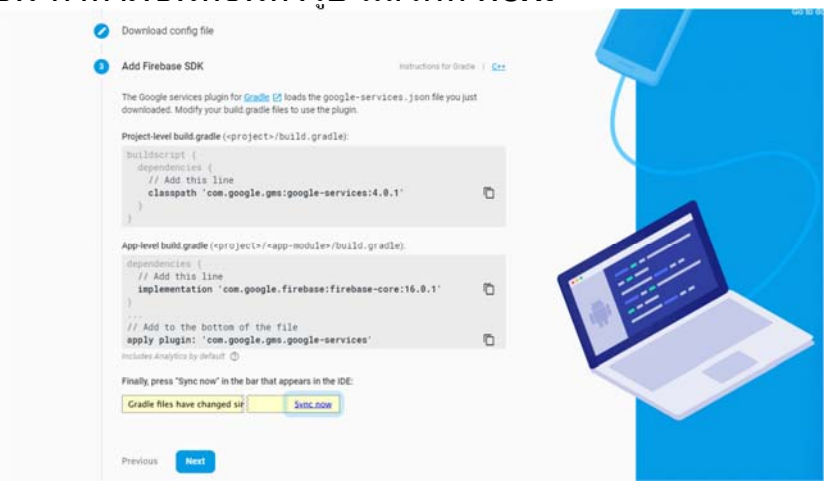
2. ใส่ชื่อ Android package เพื่อทำการลงทะเบียนแอปพลิเคชัน แล้วกด Register app



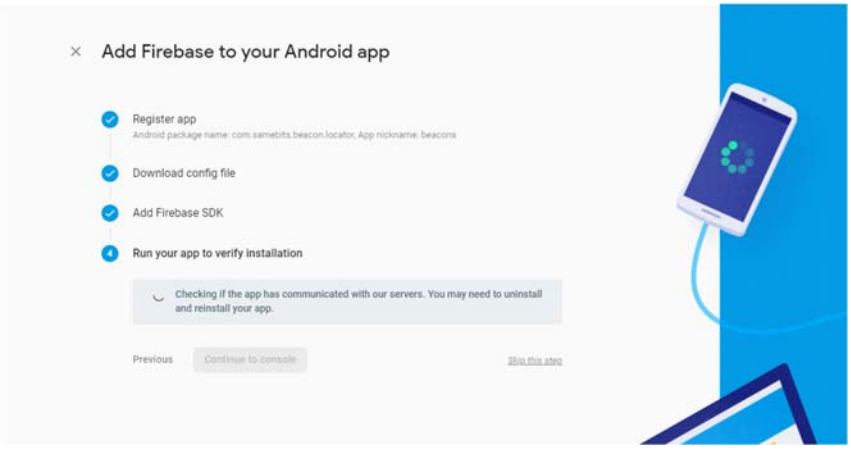
3. ดาวน์โหลดไฟล์ Config นำไปใส่ในแอปพลิเคชันตามโฟลเดอร์ดังรูป แล้วกด next



4. หลังจากนั้นทำการเพิ่ม Firebase SDK ลงในโปรเจคของแอปพลิเคชัน ทำตามขั้นตอนดังรูป แล้วกด next



5. รันแอปพลิเคชันเพื่อทำการตรวจสอบการติดตั้ง Firebase เสร็จสิ้นการติดตั้ง



ภาคผนวก ข
การทำงานในส่วนของ python

- ทำการติดตั้งผ่าน console เพื่อใช้ Google map

$pip install git+https://github.com/vgm64/gmplot

- Modules ที่ใช้ในการทำ Classification

Link	https://github.com/mpattt/mm-seniorproject/blob/master/src/Code_Beacon.ipynb
Module	Pandas
การทำงาน	ใช้ Pandas ในการโหลดไฟล์ข้อมูล excel จะได้ข้อมูลในรูปแบบตาราง(DataFrame) ที่แบ่งข้อมูลตามแถวและคอลัมน์
Module	Numpy
การทำงาน	ใช้ Numpy ในการคำนวณตัวเลขต่างๆ
Module	Sklearn
การทำงาน	ใช้ Scikit-learn ซึ่งเป็น library ใน Python ในการทำ Classification โดยกำหนดค่าแต่ละโมเดลดังนี้ ● SVC(C=1.0,cache_size=200,class_weight=None,coef0=0.0,decision_function_shape='ovo',degree=3,gamma=0.1, kernel='rbf',max_iter=-1, probability=False,random_state=None, shrinking=True,tol=0.001, verbose=False) ● RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',max_depth=None, max_features='auto',

	max_leaf_nodes=None,min_impurity_split=1e-07, min_samples_leaf=1,min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=700, n_jobs=1, oob_score=False, random_state=1,verbose=0, warm_start=False) ● KNeighborsClassifier(n_neighbors=7)
Module	Matplotlib
การทำงาน	ใช้ในการคำนวณด้านคณิตศาสตร์และพลอตกราฟสองมิติจาก array รวมทั้งบันทึกผลที่ได้ออกมาเป็นรูปภาพ

- Modules ที่ใช้ในการทำ Clustering

Link	https://github.com/mpattt/mm-seniorproject/blob/master/src/Code_poi.ipynb
Module	Pandas
การทำงาน	ใช้ Pandas ในการโหลดไฟล์ข้อมูล excel จะได้ข้อมูลในรูปแบบตาราง(DataFrame) ที่แบ่งข้อมูลตามแถวและคอลัมน์
Module	Sklearn
การทำงาน	ใช้ Scikit-learn ซึ่งเป็น library ใน Python ในการทำ Clustering โดยกำหนดค่าดังนี้ ● DBSCAN(eps=0.1, min_samples=4, metric='precomputed',algorithm ='auto')
Module	Matplotlib
การทำงาน	ใช้ในการคำนวณด้านคณิตศาสตร์และพลอตกราฟสองมิติจาก array รวมทั้งบันทึกผลที่ได้ออกมาเป็นรูปภาพ

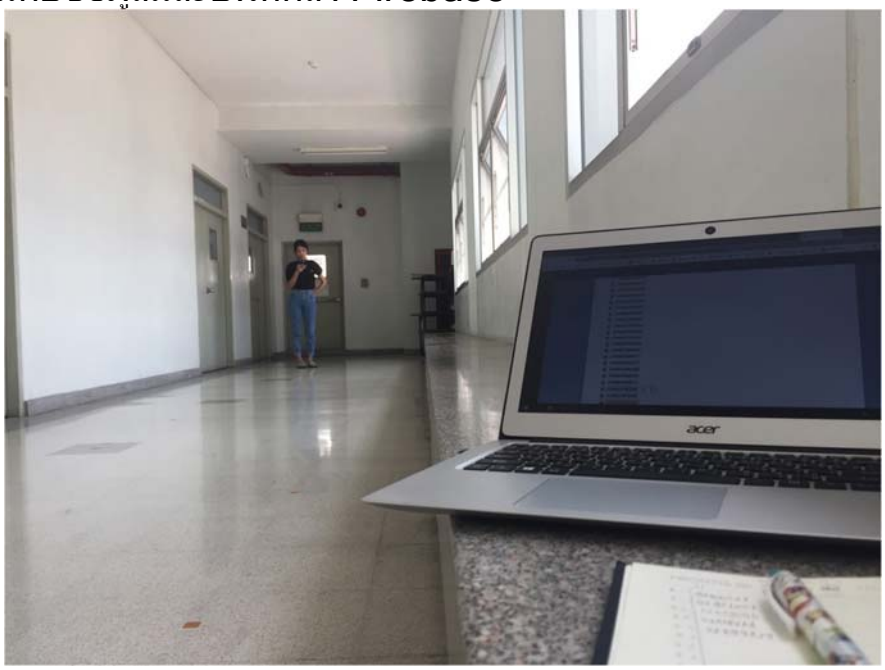
Module	Scipy
การทำงาน	ใช้ในการคำนวณระยะทางระหว่างตำแหน่ง

ภาคผนวก ค

การทำงานระหว่างการเก็บข้อมูล

- ตำแหน่งที่ติดตั้งบีคอนส์ทั้ง 3 ตัว

- การเก็บข้อมูลและบันทึกลง Firebase



ภาคผนวก ง
อัลกอริทึมที่ใช้

- อัลกอริทึมที่ใช้ในการหาตำแหน่งหยุด

Algorithm StayPoint_Detection(*P*, *distThreh*, *timeThreh*)

Input: A GPS log *P*, a distance threshold *distThreh*
and time span threshold *timeThreh*
Output: A set of stay points $SP=\{S\}$
1. $i=0$, $pointNum = |P|$; //the number of GPS points in a GPS logs
2. **while** $i < pointNum$ **do**,
3. $j:=i+1$;
4. **while** $j < pointNum$ **do**,
5. $dist$=Distance(p_i, p_j); //calculate the distance between two points
6. **if** $dist > distThreh$ **then**
7. $\Delta T=p_j.T-p_i.T$; //calculate the time span between two points
8. **if** $\Delta T>timeThreh$ **then**
9. $S.coord$=ComputMeanCoord($\{p_k \mid i<=k<=j\}$)
10 $S.arvT= p_i.T$; $S.levT=p_j.T$;
11. SP.insert(S);
12. $i:=j$; break;
13. $j:=j+1$;
14. **return** SP.