



# การทำเหมืองข้อมูลเส้นทางเดินจากข้อมูลการระบุตำแหน่ง Trajectory Data Mining Based on Location History

พิชญา ลิฬหัทธ, ภัทร โตจิตต์ และ อ.ดร.วีรยุทธ เจริญเรืองกิจ  
หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

## 01 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 the 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. However, the indoor setting is still ongoing research using several types of sensors. The second sub-problem is often categorized as trajectory mining, which 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 experiment for the first sub-problem is conducted by collecting data 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 trajectory dataset implemented by the DBSCAN method.

## 02

### องค์ความรู้ที่เกี่ยวข้อง

ระบบระบุตำแหน่งภายในอาคาร



**Beacon<sup>[1]</sup>** ตัวกวางสัญญาณจับบลูทูธจะส่งสัญญาณกวาดรอบตัวและนำสัญญาณบลูทูธที่ได้มาประมวลผลเพื่อตีความตำแหน่งที่ตั้งโดยเฉพาะ



**RSSI<sup>[2]</sup>** เป็นวิธีการวัดความเข้มสัญญาณของสัญญาณวิทยุระหว่างเซ็นเซอร์ไร้สายตัวหนึ่งไปยังเซ็นเซอร์ไร้สายอีกตัวหนึ่ง

เทคนิค Classification

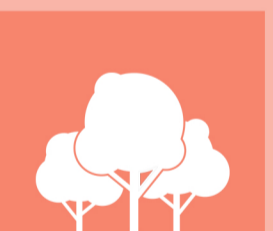
**k-Nearest Neighbor<sup>[3]</sup> (KNN)** เป็นวิธีการจำแนกจะใช้วิธีการวิเคราะห์จากข้อมูลที่ใกล้เคียงที่สุดจำนวน เค ตัวกับข้อมูลที่ต้องการจำแนกประเภทของข้อมูลหรือต้องการทำนายคลาสของข้อมูลใหม่โดยจะทำงานตามคลาสส่วนใหญ่ของข้อมูลฝึก เค ตัว



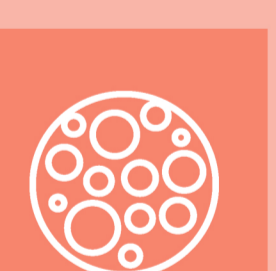
**Support Vector Classification<sup>[4]</sup> (SVC)** เป็นวิธีการจำแนกกลุ่มข้อมูลที่อาศัยระนาบการตัดสินใจที่เรียกว่า Hyperplane มาใช้ในการจำแนกกลุ่มข้อมูล



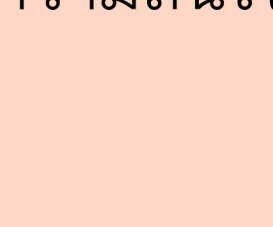
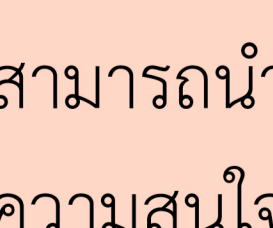
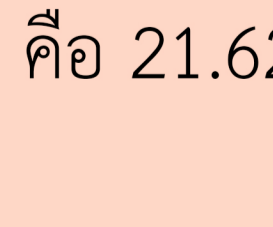
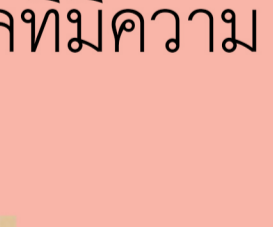
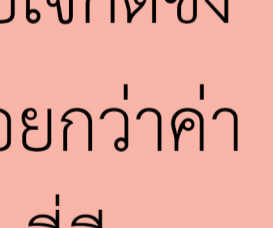
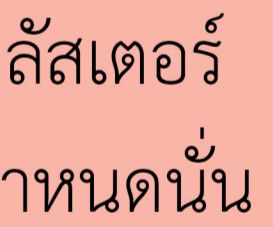
**Random Forest<sup>[5]</sup> (RF)** เป็นเทคนิคที่ทำการสุ่มเลือกแอตทริบิวต์ (พีเจอร์) ต่างๆ ออกมาเป็นหลายๆชุดด้วย และสร้างโมเดลด้วยเทคนิค Decision Tree



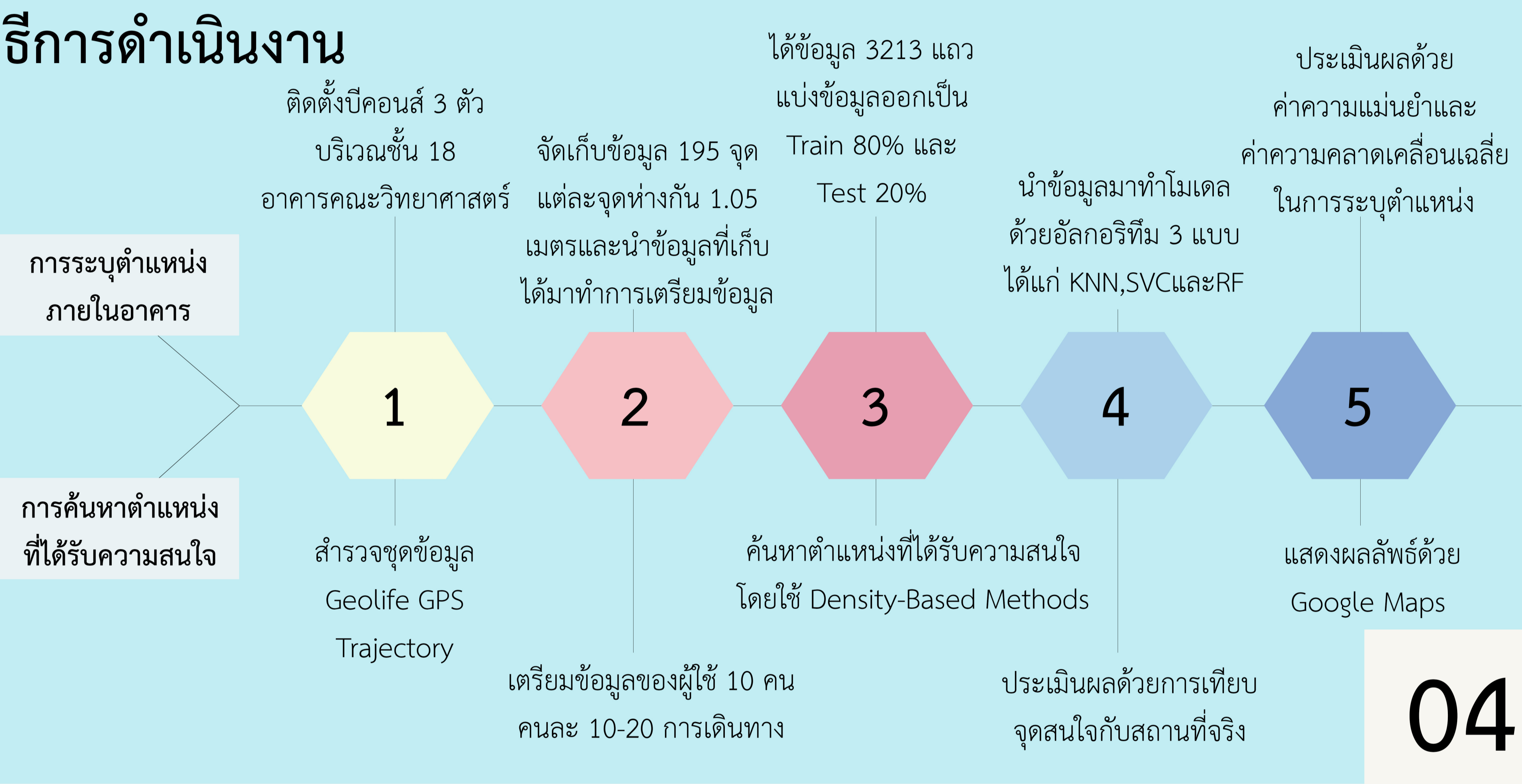
เทคนิค Clustering



**Density-Based Methods<sup>[6]</sup>** คือการแผ่ขยายขอบเขตของคลัสเตอร์ไปเรื่อยๆ ครอบคลุมถึงความหนาแน่นของอ็อบเจกต์ที่มีค่าถึงค่าที่กำหนดนั้นคือแต่ละอ็อบเจกต์ของคลัสเตอร์ใดๆ จะต้องประกอบด้วยอ็อบเจกต์ซึ่งอยู่ใกล้กันภายในรัศมีที่กำหนด (neighborhood) จำนวนไม่น้อยกว่าค่าที่กำหนดและสามารถใช้ในการกรองข้อมูลรบกวนซึ่งเป็นข้อมูลที่มีความหนาแน่นเบาบางได้



### 03 วิธีการดำเนินงาน

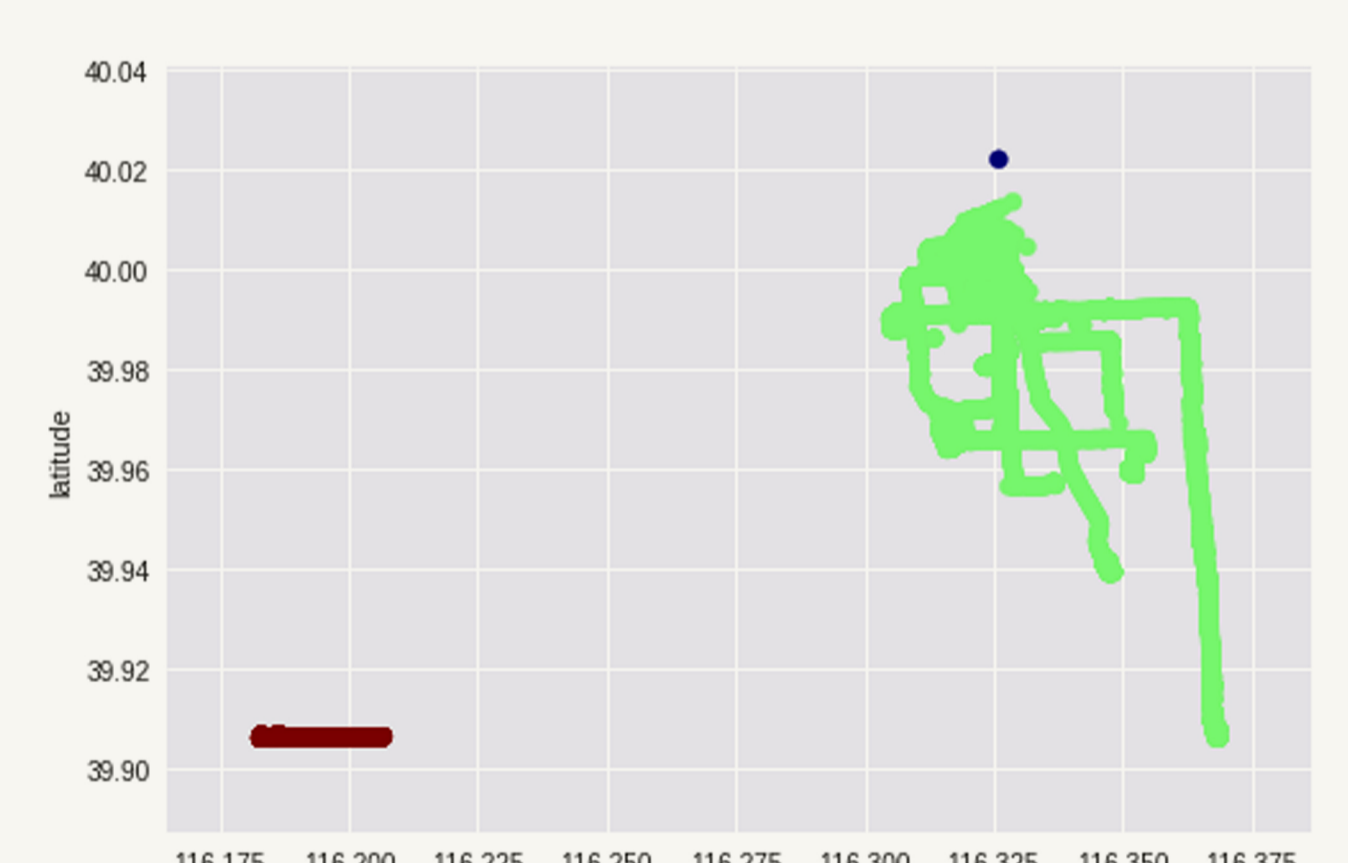
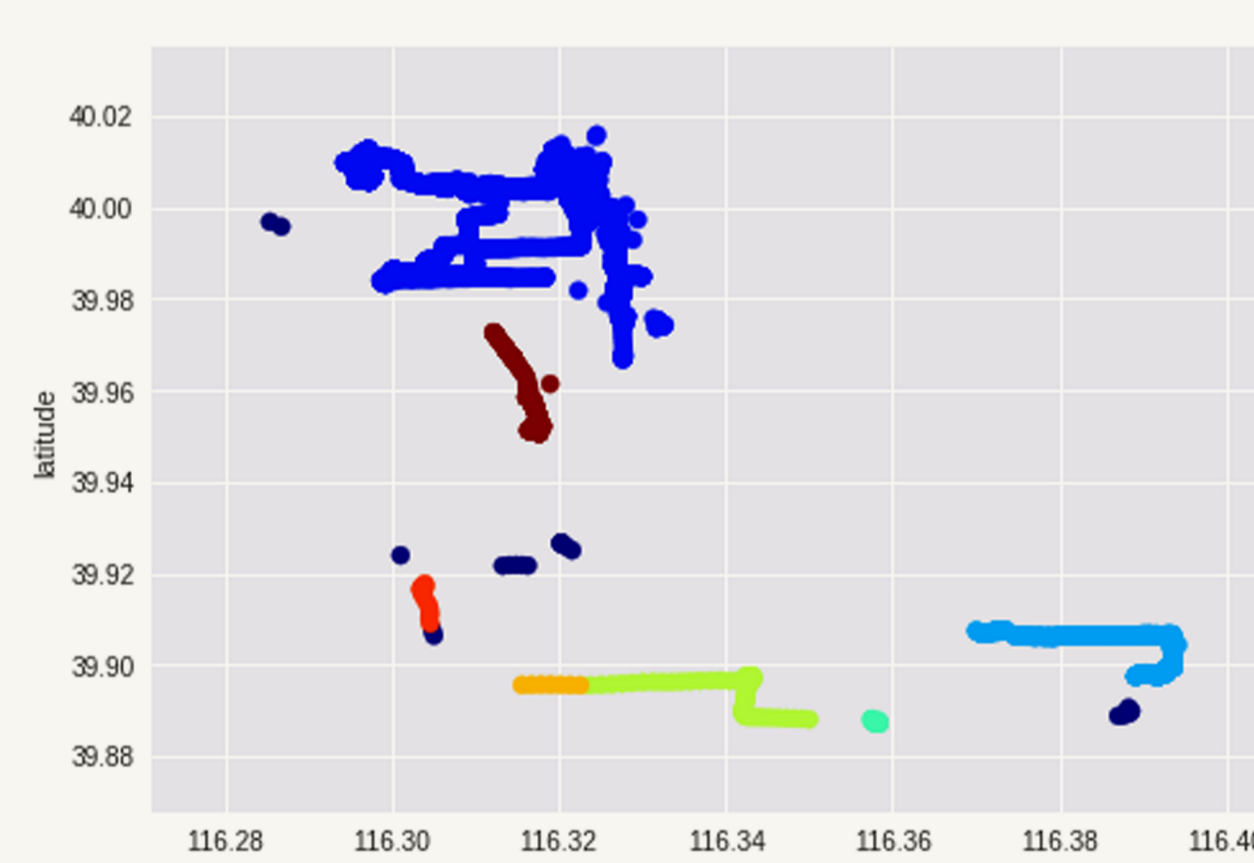
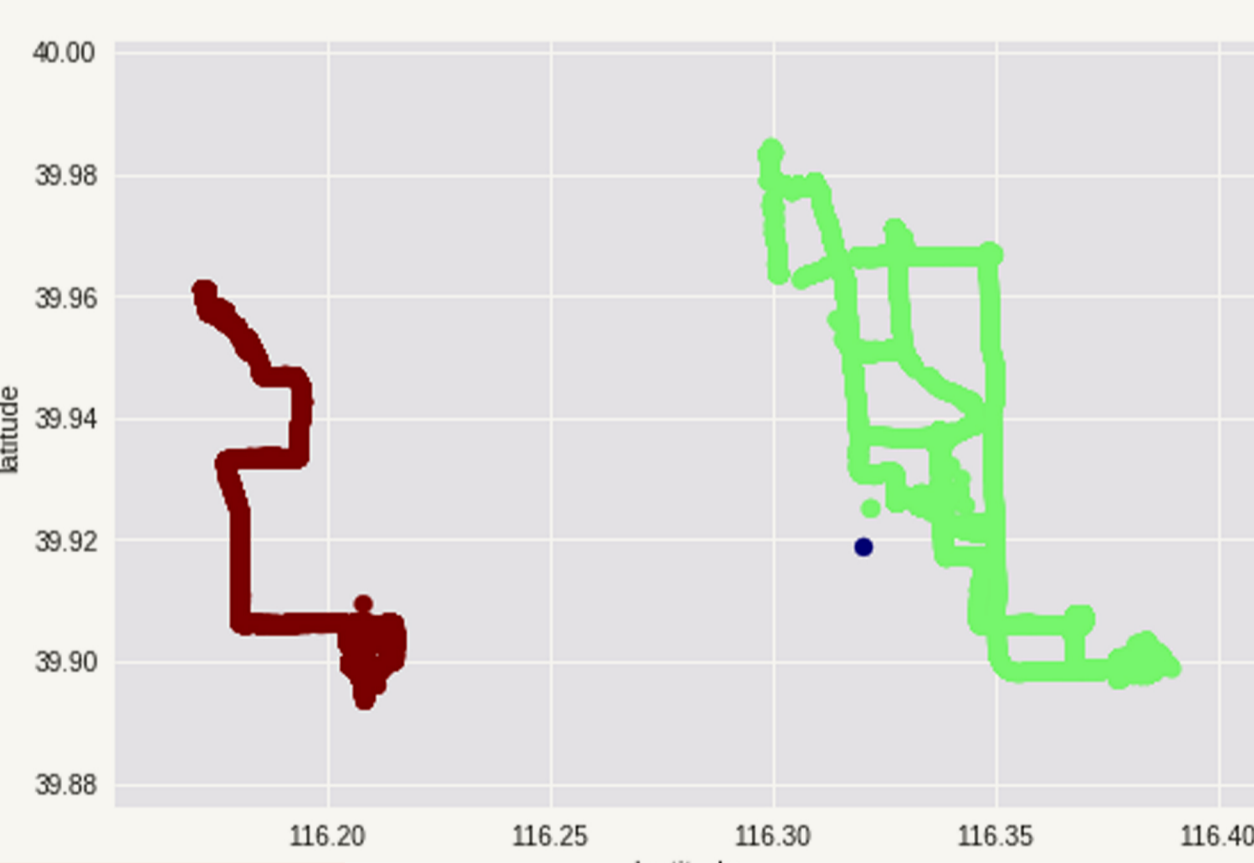


## 04

ผลการดำเนินงานในส่วนของการระบุตำแหน่งภายในอาคาร

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

ตารางผลการทดสอบความแม่นยำและความคลาดเคลื่อน



ผลการดำเนินงาน

จากตารางจะเห็นว่าค่าความแม่นยำในการระบุตำแหน่งไม่สูงมาก แต่ค่าความคลาดเคลื่อนเฉลี่ยของพิกัดไม่ห่างไกลมาก สาเหตุเนื่องจากระยะห่างของแต่ละพิกัดอยู่ที่ 1.05 เมตร

ผลการดำเนินงานในส่วนของการค้นหาตำแหน่งที่ได้รับความสนใจ

ผลการค้นหาจุดสนใจด้วย DBSCAN ด้วยระยะทางรัศมี 500 เมตร และ min\_simple = 10

## 05

### สรุปผลดำเนินงาน

- จากผลการดำเนินงานในส่วนของการระบุตำแหน่งภายในอาคาร อัลกอริทึมแบบ Support Vector Classification มีความแม่นยำในการระบุตำแหน่งสูงสุด คือ 21.62 % และมีความคลาดเคลื่อนเฉลี่ยในการระบุตำแหน่งน้อยที่สุดคือ 4.45 เมตร
- จากผลการดำเนินงานในส่วนของการค้นหาตำแหน่งที่ได้รับความสนใจ สามารถนำชุดข้อมูล Geolife GPS trajectory มาจัดกลุ่มได้ เพื่อนำไปสู่การค้นหาตำแหน่งที่ได้รับความสนใจ

## 06

### บรรณานุกรม

[1] 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>

[2] เอกสิทธิ์ พัทธวงศ์ศักดิ์. การสร้างโมเดล Ensemble แบบต่างๆ. สืบค้นเมื่อ 16 มกราคม 2562. จาก <http://dataminingtrend.com/2014/data-mining-techniques/ensemble-model/>

[3] นางสาวดิพร จันทร์กลิ่น. (2557). การปรับปรุงอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนสำหรับการจำแนกข้อมูลภาพไปโอเมตริกซ์. สืบค้นเมื่อ 21 มกราคม 2562. จาก <https://core.ac.uk/download/pdf/70943503.pdf>

[4] นายพงศกร ชีร์รัมย์. (2558). วิธีการหาค่า เค ที่เหมาะสมในการจำแนกแบบเคเนียร์เรสมอนเบอร์กับข้อมูลทางการแพทย์. สืบค้นเมื่อ 23 มกราคม 2562. จาก <http://sutir.sut.ac.th:8080/sutir/bitstream/123456789/6072/2/Fulltext.pdf>

[5] ภาควิชา สอนเสาวภาคย์และชัยย์ คุณบัว. (2552). ระบบระบุตำแหน่งวัตถุภายในอาคารโดยใช้มาตรฐาน IEEE 802.15.4. สืบค้นเมื่อ 18 พฤศจิกายน 2561. จาก [http://www.ecti-thailand.org/assets/papers/4\\_pub\\_1.pdf](http://www.ecti-thailand.org/assets/papers/4_pub_1.pdf)

[6] นายธรรมศักดิ์ เขียวนิเวศน์. (2548). การลดขนาดข้อมูลด้วยน้ำหนักความหนาแน่นเพื่อการจัดกลุ่มข้อมูลขนาดใหญ่. สืบค้นเมื่อ 29 มกราคม 2562, จาก <http://www2.it.kmutnb.ac.th/journal/pdf/vol20/ch06.pdf>