



การประยุกต์ใช้ Machine Learning กับ ระบบสัญญาณ WiFi  
ในการระบุตำแหน่งมือถือในพื้นที่ปิด  
INDOOR LOCALIZATION THROUGH WIFI FINGERPRINT  
USING MACHINE LEARNING

ปฏิพล ศรีชุ่มสิน

การประยุกต์ใช้ Machine Learning กับ ระบบสัญญาณ WiFi  
ในการระบุตำแหน่งมือถือในพื้นที่ปิด



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ  
ปีการศึกษา 2562  
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

INDOOR LOCALIZATION THROUGH WIFI FINGERPRINT  
USING MACHINE LEARNING



PATIPON SRICHOOMSIN

A Master's Project Proposal Submitted in Partial Fulfillment of the Requirements  
for the Degree of MASTER OF SCIENCE  
(Information Technology)  
Faculty of Science, Srinakharinwirot University  
2019  
Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การประยุกต์ใช้ Machine Learning กับ ระบบสัญญาณ WiFi

ในการระบุตำแหน่งมือถือในพื้นที่ปิด

ของ

ปฏิพล ศรีชุมสิน

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก ..... ประธาน  
(ผู้ช่วยศาสตราจารย์ ดร.จันตรี ผลประเสริฐ) (อาจารย์ ดร.สุทธิพงษ์ ธีชัยพงษ์)

..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.นุรีย์ วิวัฒน์วัฒนา)

## สารบัญ

	หน้า
สารบัญ .....	๑
สารบัญตาราง.....	๗
สารบัญรูปภาพ .....	๗
บทที่ 1 บทนำ.....	1
ความสำคัญ และ ที่มาของปัญหา.....	1
วัตถุประสงค์ของการวิจัย.....	3
ขอบเขตของการวิจัย .....	3
วิธีดำเนินการวิจัย.....	3
ประโยชน์ที่คาดว่าจะได้รับ.....	3
บทที่ 2 แนวคิด ทฤษฎี และ เทคโนโลยีที่เกี่ยวข้อง .....	4
Location Based Service (LBS).....	4
เทคโนโลยี Wi-Fi .....	5
หลักการทํางานของ Fingerprinting .....	6
ระยะออฟไลน์ (Offline Phase) .....	7
ระยะออนไลน์ (Online Phase) .....	7
Machine Learning .....	7
Classification .....	10
2 K Nearest Neighbors (KNN).....	11
3 AdaBoost .....	13
K Fold Cross Validation.....	14
งานวิจัยที่เกี่ยวข้อง .....	15

บทที่ 3 การพัฒนางานวิจัย .....	23
ขั้นตอนเตรียม Dataset ให้พร้อมใช้งาน.....	23
การทำ Exploratory Data Analysis และ Pre-processing.....	25
บทที่ 4 ผลการวิจัย .....	28
บทที่ 5 สรุปและแนวทางการพัฒนาที่จะทำเพิ่มเติม.....	33
สรุป 33	
การพัฒนาที่น่าจะทำเพิ่ม .....	33
บรรณานุกรม .....	34
ประวัติผู้เขียน.....	37



## สารบัญตาราง

### หน้า

ตาราง 1 แสดงพารามิเตอร์ DNN ของงานวิจัย A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting ..... 18

ตาราง 2 แสดงผลลัพธ์จาก DNN ของงานวิจัย A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting ..... 18

ตาราง 3 แสดงตัวอย่างข้อมูลใน UJIIndoorLoc Dataset ..... 24



## สารบัญรูปภาพ

### หน้า

ภาพประกอบ 1 การนำเทคโนโลยี LBS มาใช้ในการระบุและติดตามตำแหน่งลูกค้าที่มาใช้บริการ ภายในห้างสรรพสินค้า <u>Brian Friedman</u> . (2015). Location-Based Service. ....	4
ภาพประกอบ 2 Machine Learning Flowchart.....	7
ภาพประกอบ 3 Type of learning and data .....	9
ภาพประกอบ 4 Random Forest Prediction Concept.....	11
ภาพประกอบ 5 K-nearest neighbors comparative distance concept .....	12
ภาพประกอบ 6 AdaBoost weighting concept.....	13
ภาพประกอบ 7 โครงสร้าง DNN งานวิจัย Low-effort place recognition with WiFi fingerprints using deep learning .....	15
ภาพประกอบ 8 แสดงผลประสิทธิภาพของงานวิจัย Low-effort place recognition with WiFi fingerprints using deep learning .....	16
ภาพประกอบ 9 แสดงโครงสร้าง DNN ของงานวิจัย A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting .....	17
ภาพประกอบ 10 ตารางการเปรียบเทียบข้อผิดพลาดในการระบุตำแหน่งระหว่าง Multimodal กับ Wi-Fi Only [2].....	19
ภาพประกอบ 11 ตารางผลการทดสอบโดยเปรียบเทียบค่าความคลาดเคลื่อนทั้ง 3 วิธี[1].....	20
ภาพประกอบ 12 ตารางเปรียบเทียบค่าความถูกต้องเฉลี่ยในสภาพแวดล้อมที่แตกต่างกัน[3] ...	21
ภาพประกอบ 13 ตารางการเปรียบเทียบค่าความถูกต้องของ WLAN และ UMTS โดยใช้ WKNN เป็นตัวชี้วัด[4].....	22
ภาพประกอบ 14 ตารางการเปรียบเทียบค่าความถูกต้องของ K-NN, K-NN regression และ Naïve Bayes[5].....	22

ภาพประกอบ 15 แสดงพื้นที่ 3 อาคารของ University Jaume I, Spain.....	23
ภาพประกอบ 16 ตารางข้อมูล UJIIndoor .....	25
ภาพประกอบ 17 การกระจายของชุดข้อมูลการ access ของตัว WAPs .....	26
ภาพประกอบ 18 การกระจายตัวของข้อมูล RSSI .....	27
ภาพประกอบ 19 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก0และด้วย model K Nearest Neighbors มีค่า Accuracy เท่ากับ 0.9977, และ Confusion Matrix คือ.....	28
ภาพประกอบ 20 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล KNN ของตึก0 และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788, และ Confusion Matrix คือ .....	28
ภาพประกอบ 21 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล AdaBoost ของ ตึก0 .....	29
ภาพประกอบ 22 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก1 .....	29
ภาพประกอบ 23 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล KNN ของตึก1.	30
ภาพประกอบ 24 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล AdaBoost ของ ตึก1 .....	30
ภาพประกอบ 25 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก2 .....	31
ภาพประกอบ 26 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล KNN ของตึก2.	31
ภาพประกอบ 27 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล AdaBoost ของ ตึก2 .....	31

## บทที่ 1

### บทนำ

#### ความสำคัญ และ ที่มาของปัญหา

Location-Based Service (LBS) เป็นบริการที่ใช้ประโยชน์จากเทคโนโลยีไร้สายเพื่อใช้ในการระบุตำแหน่งที่อยู่ของ Smart Phone ได้อย่างแม่นยำ ซึ่งจะให้บริการออกมาในรูปแบบของ Software Application โดยที่เทคโนโลยี LBS จะเป็นการนำตำแหน่งของคนที่อยู่ภายในอาคารมาใช้ให้เกิดประโยชน์ ซึ่งในปัจจุบันมีธุรกิจให้ความสนใจ LBS ไปใช้งานกันอย่างแพร่หลาย โดยส่วนใหญ่ที่คนนิยมใช้เทคโนโลยี LBS กันมากที่สุด คือ การติดตามตำแหน่งของลูกค้าที่มาซื้อสินค้าภายในห้างสรรพสินค้า โดยการที่ผู้ประกอบการจะทำการส่งข้อมูลไปให้กับลูกค้า เพื่อให้จะให้ข้อมูลสนับสนุนการค้าเพิ่มเติม และการวิเคราะห์ตำแหน่งการติดป้ายโฆษณาในห้างสรรพสินค้าเพื่อให้ทราบว่าตำแหน่งใดควรเก็บค่าโฆษณาเท่าใดตามจำนวนผู้คนที่เดินผ่านบริเวณนั้น จะเห็นได้ว่าการระบุตำแหน่งภายในอาคารมีประโยชน์แก่ธุรกิจในปัจจุบันเป็นอย่างมาก ดังนั้น LBS ที่มีประสิทธิภาพจะต้องมีระบบระบุตำแหน่งภายในอาคารที่แม่นยำ โดยสำหรับการระบุตำแหน่งภายนอกอาคาร ซึ่งใช้บริการ Global Navigation Satellite System (GNSS) เช่น GPS ได้ถูกสร้างขึ้นและใช้งานมาแล้วกว่า 20 ปี ซึ่งเป็นสิ่งที่ขาดไม่ได้ในระบบนำทางภายนอกอาคาร อย่างไรก็ตาม ผู้คนใช้เวลาส่วนใหญ่อยู่ในอาคาร ซึ่งไม่สามารถใช้บริการ GNSS ได้ เนื่องจากสัญญาณของดาวเทียมไม่สามารถทะลุผ่านสิ่งก่อสร้างได้ และความแม่นยำของ GPS ไม่เพียงพอที่จะระบุตำแหน่งระดับห้องภายในอาคารได้โดยปัจจุบันได้มีการพัฒนาระบบระบุตำแหน่งภายในอาคารขึ้นมาหลายระบบ ซึ่งสามารถแบ่งได้เป็น 2 ประเภท คือ Infrastructure-based Systems และ Infrastructure-free Systems ซึ่งแบ่งตาม ระบบที่ต้องใช้อุปกรณ์ฮาร์ดแวร์ที่ต้องติดตั้งภายในอาคาร โดยอุปกรณ์ฮาร์ดแวร์เหล่านี้ถูกออกแบบมาเพื่อใช้ในการระบุตำแหน่งภายในอาคารโดยเฉพาะ เช่น การติดตั้งตัวส่งสัญญาณ (Beacon Indoor Tracking) ทำให้มีค่าใช้จ่ายในการติดตั้งสูง ซึ่งต่างกับระบบประเภท Infrastructure-free ซึ่งไม่มีค่าใช้จ่าย โดยจะใช้โครงสร้างเดิมภายในอาคาร เช่น เครือข่ายอินเทอร์เน็ตไร้สาย ในการให้บริการตำแหน่ง อย่างไรก็ตามระบบประเภท Infrastructure-based มักจะให้ความแม่นยำสูงกว่าเนื่องจากการใช้อุปกรณ์ฮาร์ดแวร์เพิ่มเติม อย่างไรก็ตามเมื่อ Wireless Local Area Networks (WLANs) และอุปกรณ์เคลื่อนที่ (Mobile Devices) เช่น สมาร์ทโฟน มีการใช้งานอย่างแพร่หลาย สามารถพบการทำงานได้อยู่ทุกที่ ทำให้การระบุตำแหน่งภายในอาคารแบบ Infrastructure-free โดยมีประสิทธิภาพมากขึ้น โดยวิธีที่ได้รับความนิยมมากคือการใช้สัญญาณ WiFi (WiFi Fingerprint) ซึ่งเป็นการเก็บข้อมูล

ความแรงสัญญาณ WiFi ของ Access point ในบริเวณนั้นมาใช้ในการระบุตำแหน่ง ซึ่งวิธีนี้ไม่ต้องติดตั้งอุปกรณ์ฮาร์ดแวร์เพิ่มเติม เนื่องจากเป็นการใช้โครงสร้างของระบบ WLANs ภายในอาคาร ที่ติดตั้งอยู่แล้ว อย่างไรก็ตาม WLANs ไม่ได้ถูกออกแบบมาเพื่อการระบุตำแหน่ง ดังนั้นอุปสรรคที่สำคัญของวิธีการนี้คือ สภาพแวดล้อมภายในอาคาร การกระจายของสัญญาณภายในอาคารยากแก่การทำนาย นอกจากนี้ การระบุตำแหน่งภายในอาคารโดยใช้ WLANs ผู้ใช้จะถือสมาร์ทโฟนไว้กับตัว ซึ่งการเคลื่อนไหว หรือรูปแบบการถือสมาร์ทโฟนที่แตกต่างกันก็เป็นปัจจัยที่สำคัญ ที่จะทำให้ค่าความแรงของสัญญาณ WiFi แตกต่างกัน ส่งผลให้การระบุตำแหน่งด้วยวิธี Wifi fingerprint ยังมีความคลาดเคลื่อนข้างสูง

ต่อมามีการใช้การเรียนรู้ของเครื่องจักร (Machine Learning) ซึ่งเป็นอัลกอริทึมที่สามารถเรียนรู้ข้อมูลได้ ถูกใช้ในการแก้ปัญหาประเภท การจดจำรูปแบบ (Pattern Recognition), การแบ่งแยกข้อมูล (Data Classification), การทำนาย (Prediction) และได้มีการนำการเรียนรู้ของเครื่องจักรมาใช้ในการวิเคราะห์ WiFi Fingerprint อย่างไรก็ตาม จะต้องมีการกระบวนการ Data Pre-processing ซึ่งเป็นการลดสิ่งผิดปกติในข้อมูลและเตรียมข้อมูลให้พร้อมก่อนนำไปเรียนรู้ เช่น กระบวนการลบค่าที่ไม่ถูกต้องหรือสูญหาย (Missing Values), การทำให้ข้อมูลเป็นแบบการแจกแจงปกติ (Normal Distribution), การแปลงข้อมูล (Data transformation) ซึ่งต้องอาศัยการทำ Data Visualization เพื่อให้ทราบถึงปัญหาหรือสิ่งผิดปกติในข้อมูล ทำให้สามารถปรับค่าพารามิเตอร์ (Fine-tuning) ให้เหมาะสมได้ ก่อนนำไปเรียนรู้ด้วยอัลกอริทึมการเรียนรู้ของเครื่องจักรเพื่อเรียนรู้ข้อมูล ที่ผ่านมามีงานวิจัยที่นำเทคนิคการเรียนรู้ของเครื่องจักรไปใช้ในการวิเคราะห์ความแรงของสัญญาณ WiFi เพื่อระบุตำแหน่งของโทรศัพท์มือถือที่อยู่มากมาย แต่ผลที่ได้ก็ยังไม่แม่นยำมากนักเนื่องมาจากสาเหตุหลายประการ อาทิ เช่น ความไม่แน่นอนของสัญญาณ WiFi สัญญาณรบกวนจากสภาพแวดล้อม เป็นต้น

งานวิจัยนี้นำเสนอการใช้เทคโนโลยีการเรียนรู้ของเครื่องจักรในการวิเคราะห์สัญญาณ WiFi เพื่อทำการระบุตำแหน่งของอุปกรณ์รับสัญญาณภายในพื้นที่ในอาคาร โดยเราศึกษาการใช้โมเดลการเรียนรู้ของเครื่องจักรแบบ KNN, Random Forest และ Adaboost ในการวิเคราะห์สัญญาณ WiFi เราวัดประสิทธิภาพของโมเดลโดยใช้ Confusion Matrix และ Accuracy

### วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาการนำเทคโนโลยีการเรียนรู้ของเครื่องมาใช้ในการวิเคราะห์สัญญาณ WiFi เพื่อใช้ในการระบุตำแหน่งในอาคาร
2. เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องแบบต่างๆในการเอามาประยุกต์ใช้ระบุตำแหน่งโทรศัพท์มือถือ

### ขอบเขตของการวิจัย

1. ใช้ข้อมูลสัญญาณ WiFi ในรูปของ Received Signal Strength Indicator (RSSI) ที่วัดได้ภายในอาคารมาระบุตำแหน่งของสัญญาณมือถือ
2. ใช้ UJIIndoor dataset เป็นฐานข้อมูลในการทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง
3. ใช้โมเดลการเรียนรู้ของเครื่องในการวิเคราะห์สัญญาณ WiFi เพื่อระบุตำแหน่งของโทรศัพท์มือถือเป็นชั้นในอาคาร
4. วิเคราะห์ประสิทธิภาพในรูปของ Accuracy และ Confusion Matrix

### วิธีดำเนินการวิจัย

1. ศึกษางานวิจัยที่เกี่ยวข้อง
2. ศึกษา UJIIndoor DataSet
3. ทดสอบการนำการเรียนรู้ด้วยเครื่องมาทำการวิเคราะห์สัญญาณ Wifi เพื่อทำ Classification
4. วิเคราะห์ผลการทดลองและสรุป

### ประโยชน์ที่คาดว่าจะได้รับ

เข้าใจหลักการทำงานของเทคนิคการเรียนรู้ของเครื่องและสามารถเอามาประยุกต์ใช้ในการวิเคราะห์สัญญาณ WiFi เพื่อระบุตำแหน่งของอุปกรณ์ไร้สายภายในอาคารได้อย่างมีประสิทธิภาพ

## บทที่ 2

### แนวคิด ทฤษฎี และ เทคโนโลยีที่เกี่ยวข้อง

#### Location Based Service (LBS)

Location-Based Service (LBS) เป็นบริการที่ใช้ประโยชน์จากเทคโนโลยีไร้สาย ที่ทำให้คนสามารถที่จะระบุตำแหน่งที่อยู่ของผู้ใช้ได้ ผ่านทางโทรศัพท์มือถือ ซึ่งจะให้บริการออกมาในรูปแบบของ Software Application โดยที่เทคโนโลยี LBS จะเป็นการนำตำแหน่งของคนที่อยู่ภายในอาคารมาใช้ให้เกิดประโยชน์ ซึ่งในปัจจุบันมีธุรกิจให้ความสนใจ LBS ไปใช้งานกันอย่างแพร่หลาย โดยส่วนใหญ่ที่คนนิยมใช้เทคโนโลยี LBS กันมากที่สุด คือการติดตามตำแหน่งของลูกค้าที่มาซื้อสินค้าภายในห้างสรรพสินค้า โดยการที่ผู้ประกอบการจะทำการส่งข้อมูลไปให้กับลูกค้า เพื่อที่จะให้ข้อมูลเพิ่มเติมต่างๆ



ภาพประกอบ 1 การนำเทคโนโลยี LBS มาใช้ในการระบุและติดตามตำแหน่งลูกค้าที่มาใช้บริการภายในห้างสรรพสินค้า Brian Friedman. (2015). Location-Based Service.

ที่มา: <https://blog.loopd.com/the-four-big-differences-between-beacons-and-nfc/>

ยกตัวอย่างเช่น Omni Channel ซึ่งเป็นการผสมผสานช่องทางการขายทั้งหมดของธุรกิจเข้าด้วยกัน เพื่อเป็นการอำนวยความสะดวกให้กับผู้บริโภคผ่านทุกช่องทางที่เป็นไปได้ เช่น โทรศัพท์มือถือ โทรทัศน์ คอมพิวเตอร์ วิทยุ เค็ดตาล็อก เป็นต้น ตัวอย่างเช่น Macy ซึ่งเป็นห้างสรรพสินค้าชั้นนำของ New York ได้ทำการ Track การใช้งานเว็บไซต์ของลูกค้า เมื่อลูกค้ารายหนึ่ง แสดงพฤติกรรมชัดเจนว่ามีความสนใจในรองเท้า Brand Name คู่หนึ่งแต่กลับไม่ได้ทำการซื้อผ่านเว็บไซต์ ต่อมาลูกค้ารายนั้นได้เดินทางเข้าไปในแผนกรองเท้าของห้างสรรพสินค้า Macy ซึ่ง Macy สามารถทำการ Track ลูกค้าได้ผ่านโทรศัพท์มือถือ จากแผนกรองเท้าภายในห้างสรรพสินค้าของ Macy ซึ่งมันจะสามารถนำเสนอส่วนลดในรองเท้าคู่ที่ลูกค้าสนใจอย่างอัตโนมัติ ผ่านทาง SMS หรือผ่านคำแนะนำของพนักงานในร้านค้า เพื่อกระตุ้นความสนใจ ให้ลูกค้าซื้อรองเท้าคู่นั้นในที่สุด ซึ่งวิธีการคือ ถ้าลูกค้าสนใจที่จะซื้อรองเท้า และผู้ประกอบการสามารถที่จะใช้เทคโนโลยี LBS ติดตามตำแหน่งของลูกค้าได้ โดยผู้ประกอบการจะทำการส่งข้อมูลรองเท้าไปให้กับลูกค้าผ่านช่องทางต่างๆ เพื่อที่จะเป็นการกระตุ้นให้ลูกค้าอยากที่จะซื้อสินค้าเพิ่มขึ้น

แต่การที่จะนำเทคโนโลยี LBS มาใช้งาน จำเป็นต้องมีเทคโนโลยีการระบุตำแหน่งที่แม่นยำ ซึ่งถ้าภายนอกอาคาร (Outdoor) สามารถที่จะใช้เทคโนโลยี GPS มาใช้ในการระบุตำแหน่งได้ แต่ถ้าเป็นภายในอาคาร จำเป็นต้องมีเทคโนโลยีการระบุตำแหน่งภายในอาคาร (Indoor localization) ที่มีความแม่นยำซึ่งเป็นปัญหาที่ค่อนข้างท้าทาย ว่าทำอย่างไรให้การระบุตำแหน่งนั้นมีความถูกต้อง แม่นยำ และมีประสิทธิภาพมากที่สุด

### เทคโนโลยี Wi-Fi

Wi-Fi [19] ย่อมาจาก Wireless Fidelity เป็นเทคโนโลยีเครือข่ายไร้สาย ภายใต้เทคโนโลยีการสื่อสารตามมาตรฐาน IEEE 802.11 โดยเป็นมาตรฐานที่ถูกอนุมัติให้ใช้จาก IEEE (The Institute of Electrical and Electronics Engineers) เพื่อให้อุปกรณ์คอมพิวเตอร์สามารถสื่อสารกันได้บนมาตรฐานการทำงานแบบเดียวกัน สำหรับเทคโนโลยีเครือข่ายไร้สายนี้จะใช้คลื่นความถี่ Radio Frequency (RF) และคลื่นความถี่อินฟราเรดในการรับส่งข้อมูล คลื่นความถี่วิทยุของเครือข่ายไร้สายจึงสามารถทะลุทะลวงกำแพง หรือสิ่งกีดขวางได้ทำให้การใช้งานบนเครือข่ายไร้สายมีความคล่องตัวและสะดวกสบายมากขึ้น โดยสามารถเชื่อมต่อเข้าสู่เครือข่ายได้ทุกที่ที่มีสัญญาณ ข้อมูลจะถูกส่งผ่านคลื่นวิทยุความถี่ 2.4 GHz หรือ 5 GHz

ในปัจจุบันนิยมใช้ Wi-Fi เพื่อต่อกับอินเทอร์เน็ต โดยอุปกรณ์พกพาต่างๆ เช่น คอมพิวเตอร์ส่วนบุคคล, เครื่องเล่นเกม, โทรศัพท์สมาร์ทโฟน, แท็บเล็ต เป็นต้น โดยที่สามารถเชื่อมต่อกับอินเทอร์เน็ตได้ผ่านอุปกรณ์ที่เรียกว่าแอคเซสพอยต์ (Access Point) หรือ ฮอตสปอต (Hotspot)

ซึ่งเทคโนโลยี Wi-Fi ก็ยังคงมีข้อจำกัดในการใช้งานอยู่หลายประการ ไม่ว่าจะเป็นเรื่องของความปลอดภัย การใช้งานร่วมกันของอุปกรณ์ต่างรุ่น และความยากลำบากในการทำกำไรให้กับผู้ให้บริการเครือข่าย

อย่างไรก็ตาม อัตราเร็วในการรับส่งข้อมูลผ่านอุปกรณ์ Access Point ของผู้ใช้งานแต่ละคนอาจมีค่าไม่เท่ากัน ขึ้นอยู่กับสภาพแวดล้อมในการใช้งาน และจำนวนผู้ใช้งานที่แบ่งกันรับส่งข้อมูลผ่านอุปกรณ์ Access Point ร่วมกัน นอกจากนี้ยังขึ้นอยู่กับรูปแบบในการรับส่งข้อมูลของแต่ละคนอีกด้วย

เนื่องจากสัญญาณ Wi-Fi เป็นสัญญาณที่มีอยู่แล้วภายในอาคาร เราสามารถนำอุปกรณ์ Access Point ที่มีอยู่แล้วนำมาใช้ประโยชน์ โดยการที่นำมาช่วยในการระบุตำแหน่งภายในอาคาร (Indoor Localization) วิธีการคือ ใช้ค่าความแรงของสัญญาณ Wi-Fi (RSSI) ในการที่จะเชื่อมโยงเข้ากับตำแหน่งของคนที่อยู่ภายในอาคารได้ ซึ่งวิธีการเก็บข้อมูลที่นิยมใช้กันก็คือ การเก็บข้อมูลแบบ Fingerprinting เพราะแต่ละตำแหน่ง สัญญาณ Wi-Fi มีความแรงของสัญญาณที่แตกต่างกัน ดังนั้นการใช้สัญญาณ Wi-Fi โดยการเก็บข้อมูลแบบ Fingerprinting จึงสามารถที่จะนำมาใช้ในการระบุตำแหน่งได้

### หลักการทำงานของ Fingerprinting

Fingerprinting ได้รับความนิยมเป็นอย่างมากในการนำมาใช้เพื่อระบุตำแหน่งภายในอาคาร โดยหลักการทำงานของมันคือการนำค่าความแรงของสัญญาณ Wi-Fi (RSSI of Wi-Fi) ที่วัดได้ ณ ตำแหน่งของผู้ใช้งานอยู่ขณะนั้น มาทำการเปรียบเทียบกับฐานข้อมูลที่มีการวัดค่า RSSI ไว้ก่อนหน้าแล้ว เพื่อที่จะทำการระบุตำแหน่งให้ตรงกับโหนดเป้าหมายได้อย่างถูกต้องและเกิดความคลาดเคลื่อนน้อยที่สุด ข้อดีอย่างหนึ่งของวิธีการนี้คือ ไม่จำเป็นต้องใช้อุปกรณ์ที่มีราคาสูงซึ่งเหมาะสมกับการใช้งานภายในอาคาร

Fingerprinting เป็นเทคนิคหนึ่งที่ใช้ในการระบุตำแหน่ง ซึ่งมีลักษณะการทำงานแบ่งออกเป็น 2 ขั้นตอน คือ

### ระยะออฟไลน์ (Offline Phase) เป็นระยะเก็บข้อมูล

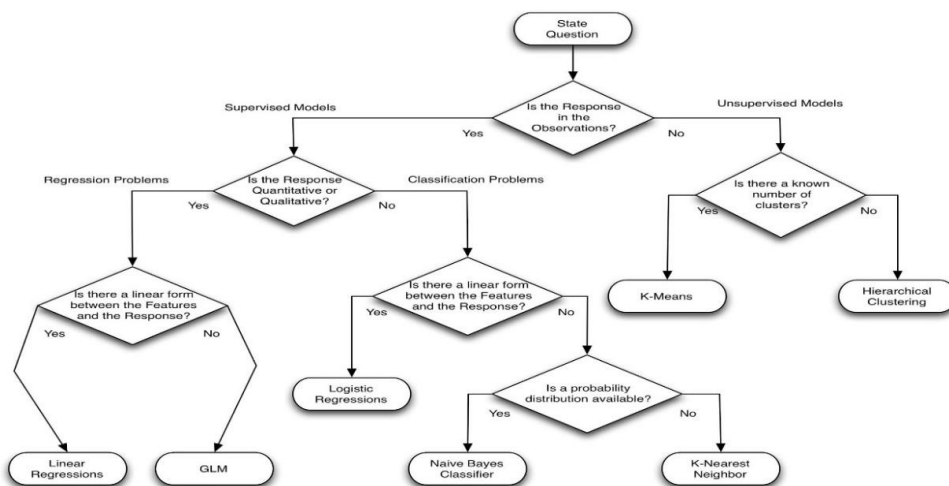
วิธีการคือ จะทำการกำหนดตำแหน่ง (Grid Point) ให้ครอบคลุมพื้นที่ที่เราต้องการ จะทำการระบุตำแหน่ง จากนั้นทำการเก็บค่าข้อมูลสัญญาณ RSSI จากหลายๆ Access Point ในบริเวณที่เราต้องการ ซึ่งเทคนิค Fingerprinting [1],[2] ข้อดีคือ ได้ค่าที่เป็นไปได้ใกล้เคียงกับการเก็บข้อมูลจริง แต่ข้อเสียคือ ใช้เวลาในการเก็บข้อมูลมาก

### ระยะออนไลน์ (Online Phase) เป็นระยะที่ใช้ในการทดสอบ

วิธีการคือ ทำการวัดค่าสัญญาณ RSSI ที่ตำแหน่ง ณ ขณะนั้นแล้วนำค่าที่ได้มาทำการเปรียบเทียบกับค่าที่เราเก็บไว้ในระยะเก็บข้อมูล เมื่อทำการเปรียบเทียบแล้วจะได้ตำแหน่งของผู้ใช้ออกมา

ข้อเสียของวิธีนี้คือ เมื่อสภาพแวดล้อมของสถานที่เปลี่ยนแปลงไป เช่น การเคลื่อนย้ายเฟอร์นิเจอร์ภายในห้องจะทำให้ Fingerprint ของตำแหน่งนั้นเปลี่ยนแปลงไป ซึ่งจำเป็นต้องมีการเก็บข้อมูลใหม่ทุกครั้ง และมีความไม่แน่นอนของสัญญาณ เช่น เมื่อมีผู้คนเดินในพื้นที่จะทำให้กระทบต่อความแรงสัญญาณ ทำให้ข้อมูลที่เก็บได้ อาจไม่แม่นยำเมื่อนำมาใช้งานจริง

## Machine Learning



ภาพประกอบ 2 Machine Learning Flowchart

การเรียนรู้ด้วยเครื่องจักร หรือ Machine Learning คือศาสตร์ที่ยืดหยุ่นและสามารถปรับใช้ให้เข้ากับลักษณะของชุดข้อมูลที่น่ามาวิเคราะห์และสร้างกระบวนการทำนายผลตามลักษณะโมเดลซึ่งข้อมูลที่จะนำไปทำนายผลตามจุดประสงค์ในการวิเคราะห์ข้อมูลก็ต้องถูกปรับแต่งให้ตรงคุณลักษณะตามเป้าหมายนั้นก่อนด้วย เช่น การวิเคราะห์ข้อมูลหุ้นเพื่อทำนายผลกระทบทางเศรษฐกิจ การวิเคราะห์ข้อมูลยอดlikeของโฆษณาเพื่อกำหนดทิศทางการตลาดจากภาพที่3กระบวนการเรียนรู้ด้วยเครื่องจักรเป็นการศึกษาความสัมพันธ์ของข้อมูลแล้วนำไปเข้ากระบวนการคัดกรองตามแต่ละโมเดลที่มีคุณสมบัติเฉพาะเพื่อทำการวิเคราะห์แล้วส่งผลวิเคราะห์กลับมาใช้ตามวัตถุประสงค์ที่ตั้งไว้ซึ่งการวิเคราะห์นั้นไม่จำเป็นต้องทำแค่ครั้งเดียว โดยผลการวิเคราะห์ที่ได้สามารถบอกความเป็นไปได้ตามวัตถุประสงค์ที่เราตั้งไว้ในการทดสอบซึ่งมันจะชี้ทิศทางของกระบวนการดำเนินงานในขั้นตอนต่อไปได้นั่นเอง เริ่มต้นที่ข้อมูลที่จะนำมาใช้ทำการเรียนรู้ด้วยเครื่องจักรนั้นเป็นข้อมูลที่เก็บรวบรวมมาโดยมีวัตถุประสงค์ในการรวบรวมไว้ล่วงหน้าก่อนที่ข้อมูลจะเกิดจริง หรือเป็นข้อมูลที่เกิดขึ้นจากเหตุการณ์จริงที่ผ่านมาแล้ว โดยหากเป็นข้อมูลที่เกิดจากเหตุการณ์จริงที่ผ่านมาแล้วก็ต้องดูว่าข้อมูลถูกแบ่งออกเป็นกลุ่มหรือ cluster หรือเปล่า เช่น ข้อมูลนักเรียนชาย นักเรียนหญิง ถ้าข้อมูลนั้นสามารถจำแนกเป็นกลุ่มได้ชัดเจนก็นำไปสู่กระบวนการหาค่าเฉลี่ยรวม หรือ K-Means โดย K คือจำนวนกลุ่มที่นำมาคิดค่าเฉลี่ยรวมแต่หากข้อมูลยังไม่ถูกแบ่งกลุ่มก็ต้องทำการแบ่งกลุ่มตามลำดับขั้นของข้อมูลถ้าข้อมูลที่จะนำมาเรียนรู้ด้วยเครื่องจักรนั้นเป็นข้อมูลที่รวบรวมเพื่อวัตถุประสงค์เฉพาะที่บ่งบอกข้อมูลทาง ปริมาณ และคุณภาพ ได้ไหม โดยหากไม่มีข้อมูลส่วนที่สามารถบ่งบอกทางปริมาณและคุณภาพได้ก็ต้องดูต่อว่าชุดข้อมูลมีคุณลักษณะเชิงเส้นหรือ linear form ด้วยหรือเปล่าซึ่งหากไม่มีก็ต้องดูว่าข้อมูลมีคุณลักษณะทางความน่าจะเป็นหรือ probability distribution หรือเปล่าโดยหากไม่มีก็หมายความว่าข้อมูลแต่ละrecordนั้นเป็นอิสระต่อกันดังนั้นจะใช้โมเดล KNN แต่ถ้าข้อมูลมีคุณสมบัติทางความน่าจะเป็นก็จะใช้โมเดล Naive Bays classification ส่วนกรณีชุดข้อมูลมีคุณสมบัติเชิงเส้นด้วยก็จะใช้โมเดล Logistic regression แต่หากข้อมูลมีคุณลักษณะทาง ปริมาณ และ คุณภาพ ก็เป็นปัญหาจำพวก Regression โดยต้องดูต่อว่าข้อมูลมีความสัมพันธ์เชิงเส้นระหว่างกันไหมโดยหากมีก็ใช้โมเดล Linear regression แต่หากไม่มีก็ใช้ โมเดล Generalized Linear model หรือ GLM และหลังจากได้เรียนรู้รูปแบบการวิเคราะห์ชุดข้อมูลก่อนจะนำชุดข้อมูลไปทำการวิเคราะห์ด้วยการเรียนรู้ด้วยเครื่องจักรแล้วก็ต้องมาเรียนรู้ถึงความแตกต่างของชุดข้อมูลแต่ละแบบดังต่อไปนี้

ชนิดของชุดข้อมูล แบ่งได้สองลักษณะใหญ่ๆ ที่สามารถจำแนกวิธีการเรียนรู้ด้วยเครื่องจักรจะสามารถนำไปใช้ได้โดยแต่ละชนิดข้อมูลก็สามารถแบ่งได้อีกสองลักษณะย่อยๆ โดยดูจากภาพประกอบ3

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

ภาพประกอบ 3 Type of learning and data

ที่มา : <https://bit.ly/2LvaxQA>

จากภาพที่ 3 จะแสดงให้เห็นได้ถึงความสัมพันธ์ของชนิดข้อมูลและรูปแบบวิธีการที่จะนำไปใช้ในกระบวนการเรียนรู้ด้วยเครื่องจักรโดยชนิดข้อมูลแบ่งได้สองลักษณะคือ 1.ข้อมูลแบบต่อเนื่อง 2.ข้อมูลแบบไม่ต่อเนื่อง และจะสามารถแบ่งชนิดการเรียนรู้ได้สองแบบ ดังนี้

1. Supervised Learning เทคนิคการเรียนรู้ของเครื่องประเภทการเรียนรู้แบบมีผู้สอน มีหลักการเรียนรู้จากข้อมูลนำเข้าที่จัดเตรียมไว้ก่อนหน้า ซึ่งประกอบด้วยข้อมูลคุณลักษณะ (Feature) และข้อมูลคำตอบ (Label) จากนั้นจึงทำการทำนายผลลัพธ์ (Output Data) จากสิ่งที่ได้เรียนรู้ด้วยลักษณะ ในบางครั้งการทำนายนี้ทำให้ในขั้นตอนการเรียนรู้ต้องอาศัยข้อมูลจำนวนค่อนข้างมากเพื่อให้ได้ผลการทำนายที่แม่นยำมากขึ้น เทคนิคการเรียนรู้แบบมีผู้สอนที่นิยม ได้แก่ การจำแนกประเภท (Classification) และการถดถอย (Regression) โดยข้อมูลที่ใช้สำหรับการเรียนรู้จะต้องมีคุณสมบัติที่สนับสนุนการเรียนรู้และการแก้ปัญหาเรียกว่า Feature Vector ในขั้นตอนการเรียนรู้ของเครื่อง (Training) นั้นจะดำเนินการโดยใช้อัลกอริทึม (Algorithm) ที่เหมาะสมกับลักษณะของปัญหาที่ต้องการแก้ไข จนกระทั่งได้โมเดลที่ดีที่สุดสำหรับสร้างเป็นแบบจำลองดัง

แสดง หลังจากขั้นตอนการเรียนรู้ แบบจำลองที่ได้จะถูกนำมาประยุกต์ใช้งาน โดยการนำชุดข้อมูล อีกหนึ่งชุด เรียกว่า Test data เข้าสู่แบบจำลอง เพื่อทำนายผลลัพธ์

2. Unsupervised Learning เทคนิคการเรียนรู้ของเครื่องประเภทการเรียนรู้แบบไม่มีผู้สอน มีหลักการเรียนรู้จากข้อมูลนำเข้าที่มีเพียงข้อมูลคุณลักษณะเท่านั้น โดยการเรียนรู้และการทำนายผลนั้นจะอาศัยการจำแนกและสร้างรูปแบบจากข้อมูลนำเข้าโดยจะไม่มีกำหนดผลลัพธ์ ที่ได้ล่วงหน้า เทคนิคที่นิยม ได้แก่ การลดมิติของข้อมูล (Dimensionality Reduction) และการจัดกลุ่ม(Clustering)

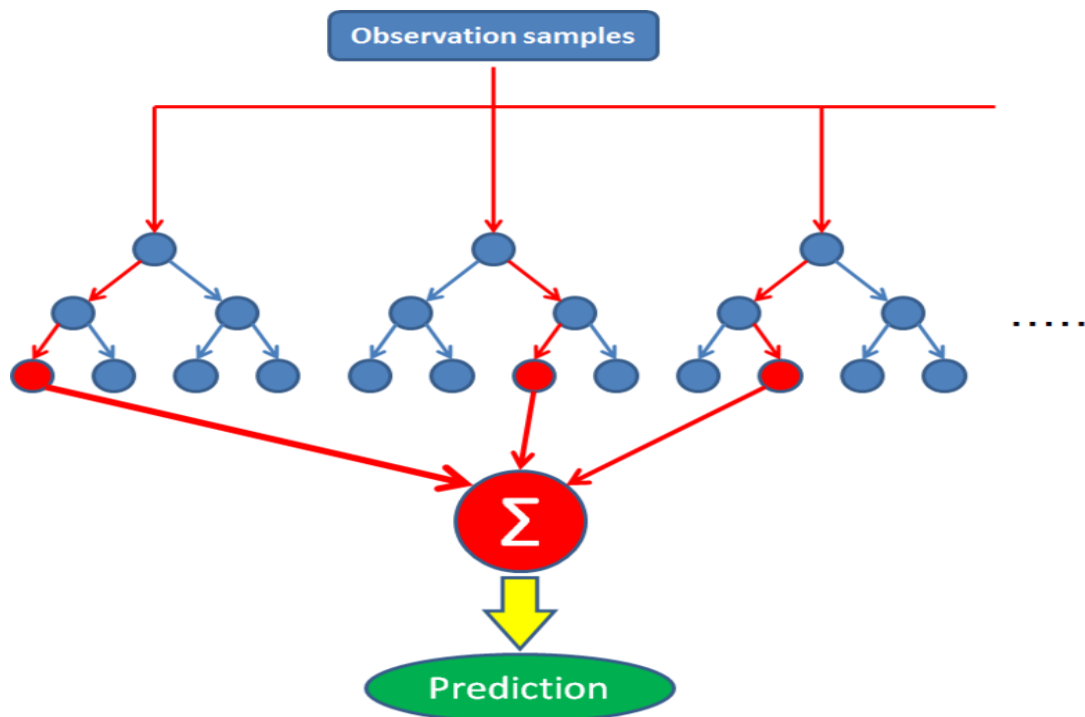
### Classification

เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่องแบบมีผู้สอน โดยกระบวนการเรียนรู้ นั้น จำเป็นต้องมีข้อมูลคำตอบ ซึ่งจะแบ่งข้อมูลคำตอบออกเป็น Class ต่างๆ สำหรับหลักการ ทำงานของแบบจำลองจะแบ่งข้อมูลออกเป็น 2 ส่วน คือชุดข้อมูลการเรียนรู้และชุดข้อมูลทดสอบซึ่งใน งานวิจัยนี้แบ่งข้อมูลดังกล่าวในอัตราส่วน 80 : 20 หรือตามความเหมาะสม จากนั้นนำข้อมูลชุด การเรียนรู้ป้อนเข้าสู่ระบบเพื่อเข้าสู่ขั้นตอนการเรียนรู้โดยใช้อัลกอริทึมที่กำหนด จนกระทั่งได้ ผลลัพธ์ออกมาเป็นแบบจำลองการทำนาย จากนั้นนำข้อมูลชุดทดสอบซึ่งเป็นข้อมูลที่เตรียมไว้ สำหรับทดสอบประสิทธิภาพการทำงานของแบบจำลองมาใช้ในการทำนายผลและพิจารณาผลลัพธ์ที่ได้

รายละเอียดแบบจำลองที่เอามาใช้จำแนก

#### 1 Random Forest

เป็นการแบ่งข้อมูลออกจากกันเป็นกลุ่มโดยจับกลุ่มข้อมูลตามคุณลักษณะที่ เกี่ยวเนื่องกันไว้ด้วยกันโดยใช้เทคนิค Decision Tree ตามจำนวนกลุ่มที่ต้องการแบ่งแยกโดย เอา ข้อมูลที่มีคุณลักษณะคล้ายกันหรือเหมือนกันมารวมกัน และหากมีต้นไม้หลายๆต้นจะเรียกว่าป่า หรือ Forest โดยการทำการสุ่มค่า ณ จุดที่ Decision Tree มีความลึกหรือความสูงมากที่สุดจะ เรียกว่าค่า K หรือ node สีแดงดังรูป  $K = 3$  แล้วนำมาหาผลรวมแล้วเอาค่าที่สุ่มมาเปรียบเทียบ เพื่อหาค่าเฉลี่ยที่ใกล้เคียงแล้วระบุชุดข้อมูลจากต้นไม้ต้นที่มีค่าเฉลี่ยใกล้เคียงค่าเฉลี่ยมากที่สุดนั่น คือสาเหตุทำไมจึงเรียก Random Forest



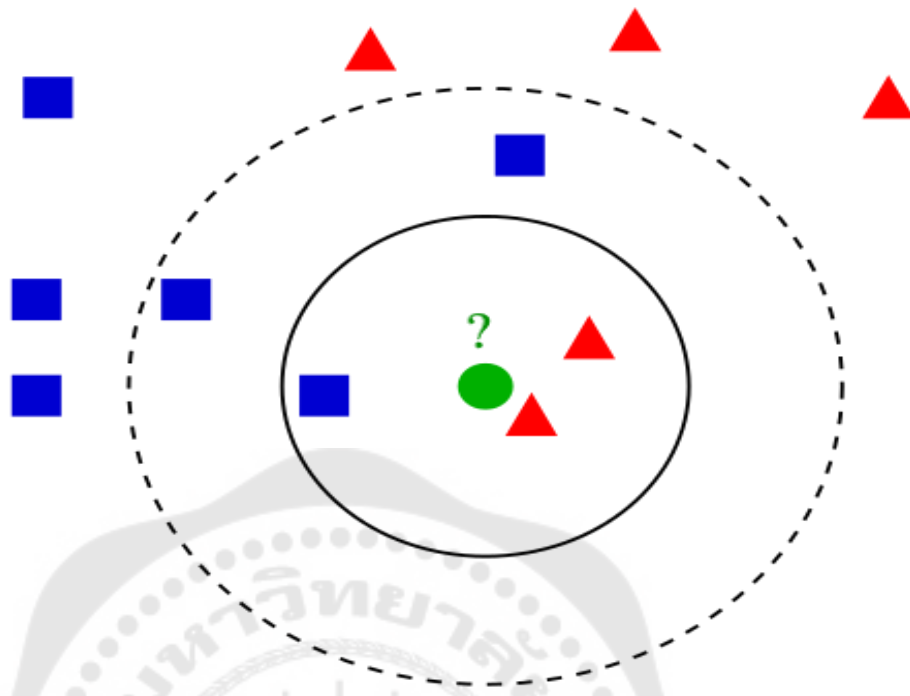
ภาพประกอบ 4 Random Forest Prediction Concept

ที่มา : <http://tjo-en.hatenablog.com/entry/2015/06/04/190000>

จากภาพที่ 4 เป็นการจำแนกข้อมูลออกเป็นจำนวน  $n$  ชุด โดยแต่ละชุดข้อมูลที่จำแนกออกมาได้อยู่ในรูปแบบ Decision Tree แล้วจะนำค่าความลึกที่  $K$  ในแต่ละต้นไม้มารวมกันแล้วหาค่าเฉลี่ยก่อนนำไปเข้าสู่กระบวนการทำนายผล

## 2 K Nearest Neighbors (KNN)

เป็นอัลกอริทึมที่พัฒนาโดยมีคุณลักษณะตรงตามตัวแบบจาก Euclidean Distance formula โดย จะทำการค้นหาข้อมูลตัวที่อยู่ใกล้กับข้อมูลหลักที่ใช้เทียบเคียงมากที่สุดออกมาโดยจะมีการค้นหาจำนวนข้อมูลเป็นจำนวน  $K$  ข้อมูลโดยจะเพิ่มระยะการค้นหามากขึ้นๆ จนกว่าจะครบจำนวน  $K$  ข้อมูล จากนั้นจะระบุกลุ่มข้อมูลที่ต้องการที่อยู่ใกล้ตัวเทียบเคียงออกมา เพราะบางครั้งอาจเจอข้อมูลที่ไม่สามารถจัดเข้ากลุ่มข้อมูลที่ต้องการได้เพราะมีคุณลักษณะที่แตกต่างไปจากคุณลักษณะของข้อมูลที่เราเจอมาส่วนใหญ่



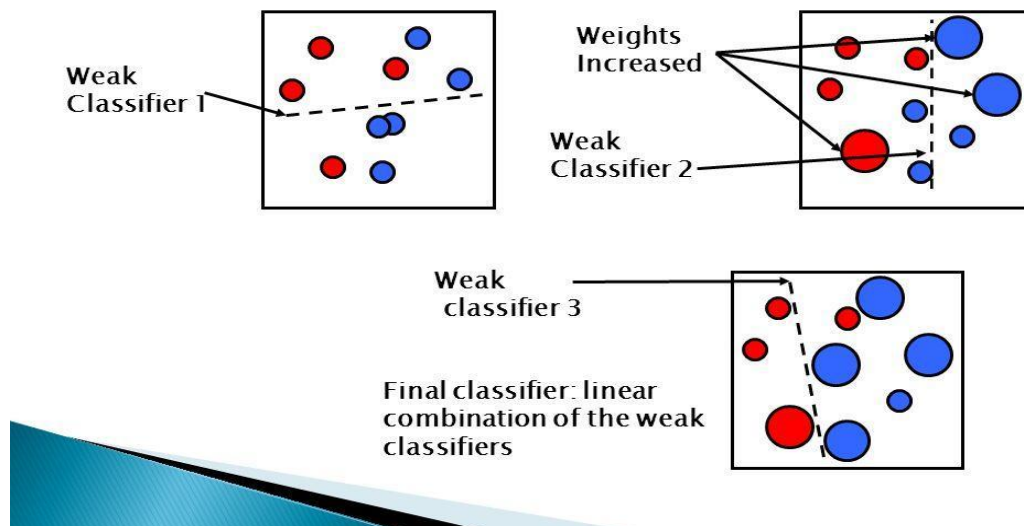
ภาพประกอบ 5 K-nearest neighbors comparative distance concept

ที่มา : [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)

จากภาพที่ 5 เป็นการเลือกกลุ่มข้อมูลโดยใช้ข้อมูลวงกลมสีเขียวเป็นตัวหลักในการเทียบเคียงระยะห่าง โดย ข้อมูลที่ใกล้ที่สุดอันดับแรกและข้อมูลที่ใกล้ที่สุดอันดับสองคือ รูปสามเหลี่ยม และ ข้อมูลที่ใกล้ที่สุดอันดับที่สามคือ ข้อมูลรูปสี่เหลี่ยม ดังนั้นถ้าวัดโดยจำนวน K คือ 3 จะได้กลุ่มข้อมูลสามเหลี่ยมเป็นผลลัพธ์ของการทำนายค่า กรณีต่อมา ข้อมูลที่อยู่ใกล้ที่สุดเป็นอันดับที่ 4 และ ข้อมูลที่อยู่ใกล้ที่สุดเป็นอันดับที่ 5 คือข้อมูลรูปสี่เหลี่ยมจากนั้นจะได้ผลลัพธ์การทำนายค่าเป็นกลุ่มข้อมูลรูปสี่เหลี่ยม โดยผลการทำนายที่ผลลัพธ์เป็นชุดข้อมูลรูปสี่เหลี่ยมมีความน่าเชื่อถือมากกว่าเพราะจำนวนข้อมูลที่ใกล้เคียงกับข้อมูลที่เอาไว้เปรียบเทียบที่มากกว่าแบบชุดข้อมูลรูปสามเหลี่ยมดังนั้นหากพิจารณาโดยค่า K คือ 5 ก็จะได้ชุดข้อมูลสี่เหลี่ยมเป็นผลลัพธ์

## 3 AdaBoost

## AdaBoost (illustrated)



ภาพประกอบ 6 AdaBoost weighting concept

ที่มา : <http://bit.ly/2ysSHpG>

จากภาพที่ 6 จะเห็นได้ว่า AdaBoost เป็น Algorithm ที่พัฒนาจากแนวคิดที่เกี่ยวกับการนำความน่าจะเป็นมาใช้เป็นตัวชี้วัดโดยจะพยายามทำการเพิ่มค่าของความน่าจะเป็นของชุดข้อมูลที่มีค่าความน่าจะเป็นน้อยให้มีค่าเพิ่มขึ้นแล้วค่อยทำการ classification โดย หากสามารถทำการเพิ่มค่าความน่าจะเป็นของชุดข้อมูลที่น่ามาทำนายค่าได้แล้วจึงทำการ classification ต่ออีกในกรณี ที่ความน่าจะเป็นที่หาออกมาได้มีค่าสูงกว่าค่าความน่าจะเป็นรอบก่อนหน้า และ หยุด ในกรณีที่ค่าความน่าจะเป็นที่หาออกมาได้มีค่าต่ำกว่าค่าความน่าจะเป็นที่หาได้ในรอบก่อนหน้าและนำผลที่ได้ไปทำการทำนายต่อไป

แนวคิดค่าน้ำหนักหรือค่าความน่าจะเป็นที่ใช้ในการชี้วัดข้อมูลของ AdaBoost

1.สมการที่ 1

$$f(x) = \sum_{t=1}^T t (\alpha_t h_t(x)) = \text{strong model (1)}$$

= ผลรวมของผลคูณของน้ำหนักกับ weak model(h)

ที่ชุดข้อมูล X ณ ข้อมูลที่ t

โดยหลังจาก classifier แล้วหากยังได้ weak model ก็จะทำกรเพิ่มน้ำหนักหรือค่าสมการที่ 2 โดยที่

$$\alpha_t = \left(\frac{1}{2}\right) \ln(\epsilon_t)$$

$\epsilon_t$  คือ Error ณ ข้อมูลที่ t,  $\alpha_t$  คือค่าน้ำหนักของข้อมูล ณ เวลาที่ t

จากภาพที่ 6 จะเห็นกลุ่มข้อมูลที่มีปริมาณน้อยกว่าในแต่ละกลุ่มจากการ classification ไปแล้วจึงทำการเพิ่มน้ำหนักค่าความน่าจะเป็นของชุดข้อมูลแล้วจึงทำการ classification อีก จากงานวิจัย[1] และ Data Set ที่ไม่มีข้อมูลทางความน่าจะเป็นให้นำมาใช้โดยตรงจึงอาจต้องเพิ่ม column ค่าความน่าจะเป็น โดยให้มีค่า 0 กับ record ที่มีค่า access point เป็นค่า 100 ทั้งหมด และ ให้มีค่าเป็น 1 กับ record ที่มีค่า access point ไม่ใช่ 100 ทุกตัว จากนั้นจึงใช้ AdaBoost model ทำการ classification

### K Fold Cross Validation

Cross Validation คือ วิธีการที่นำมาใช้เพื่อตรวจสอบ Machine Learning model ว่ามีประสิทธิภาพมากน้อยเพียงใดโดยชุดข้อมูลที่จะนำมาตรวจสอบต้องเป็นชุดข้อมูลที่ไม่ได้ผ่านกระบวนการ Machine Learning มาก่อน โดย กระบวนการ Cross Validation มีสองส่วนดังนี้

1. Crossvalidation training set
2. Crossvalidation testing set หรือ Validation set

กระบวนการทำ Cross Validation เริ่มโดยทำการ train Machine Learning model ด้วย cross validation training set แล้วทำการตรวจสอบ Machine Learning model ด้วยค่า Validation set จากนั้นนำค่าที่ได้จากทั้งสองการตรวจสอบมาเปรียบเทียบกันและมีเทคนิคที่ช่วยในการลดค่าความเบี่ยงเบนด้วยการทำ Cross Validation หลายๆครั้งโดยแต่ละครั้งใช้ Training set และ Testing set ที่ต่างกันแล้วจึงนำผลทั้งหมดมาหาค่าเฉลี่ยก่อนนำมาใช้และ K Fold Cross Validation คือ Cross Validation รูปแบบหนึ่งที่ยอมรับใช้โดยแบ่งข้อมูลออกเป็นหลาย K ชุดแล้วนำ

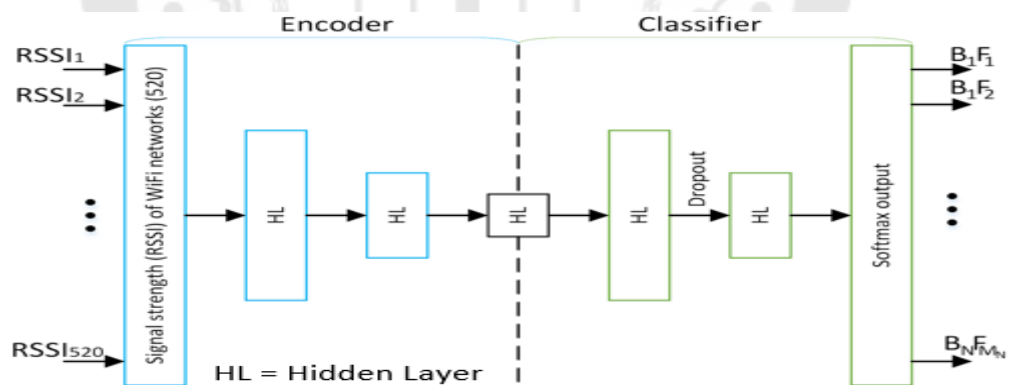
ข้อมูลจากชุดที่ 1 ถึงชุดที่ K-1 ไปทำการ Cross Validation จากนั้นนำค่าที่ได้จากการ Cross Validation ทั้งหมดมาหาค่าเฉลี่ยอีกที

### งานวิจัยที่เกี่ยวข้อง

จากการที่ได้ศึกษางานวิจัยที่เกี่ยวข้องกับการระบุตำแหน่งภายในอาคาร ผู้จัดทำจะขอยกตัวอย่างมาบางส่วน เพื่อศึกษาและนำมาปรับใช้กับโครงงานได้ดังต่อไปนี้

งานวิจัยเรื่อง UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problems ได้เสนอ และจัดทำฐานข้อมูล UJIIndoorLoc เพื่อให้เป็นฐานข้อมูลสาธารณะที่ใช้งานวิจัยทางด้านการระบุตำแหน่งภายในอาคาร โดยใช้ WiFi Fingerprint และเพื่อเป็นฐานข้อมูลกลางในการเปรียบเทียบประสิทธิภาพระหว่างงานวิจัยต่างๆ

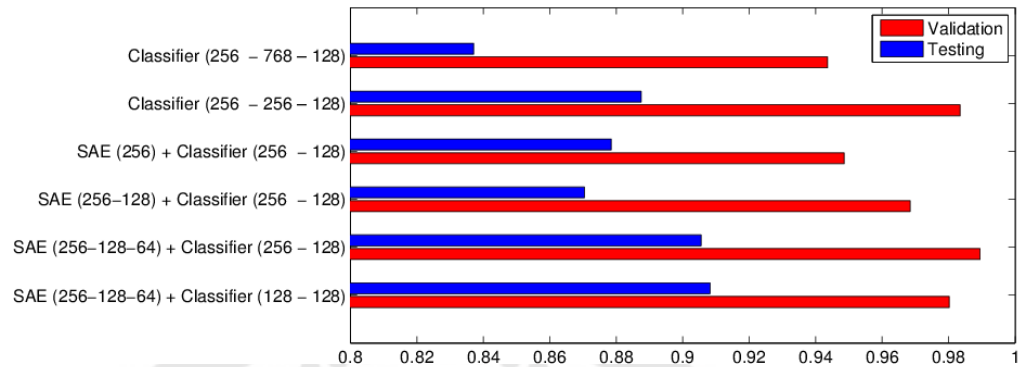
งานวิจัยเรื่อง Low-effort place recognition with WiFi fingerprints using deep learning ได้เสนอโครงสร้างของ DNN ซึ่งประกอบด้วย Stacked Autoencoder และ Classifier เพื่อใช้ในการแบ่งแยกอาคารและชั้น โดยมีโครงสร้างดังรูปที่ 8



ภาพประกอบ 7 โครงสร้าง DNN งานวิจัย Low-effort place recognition with WiFi fingerprints using deep learning

ที่มา : [ออนไลน์] <https://arxiv.org/abs/1611.02049>

ประกอบด้วย 2 ส่วนคือ Encoder และ Classifier ถูกใช้ในการแบ่งแยกอาคารและชั้น มีอินพุต คือ WiFi Fingerprint จำนวน 520 ค่า ซึ่งจะถูกย่อยข้อมูลเหลือ 64 ค่าและจะถูกส่งต่อไปยัง Classifier เพื่อแบ่งแยกชั้นและอาคาร

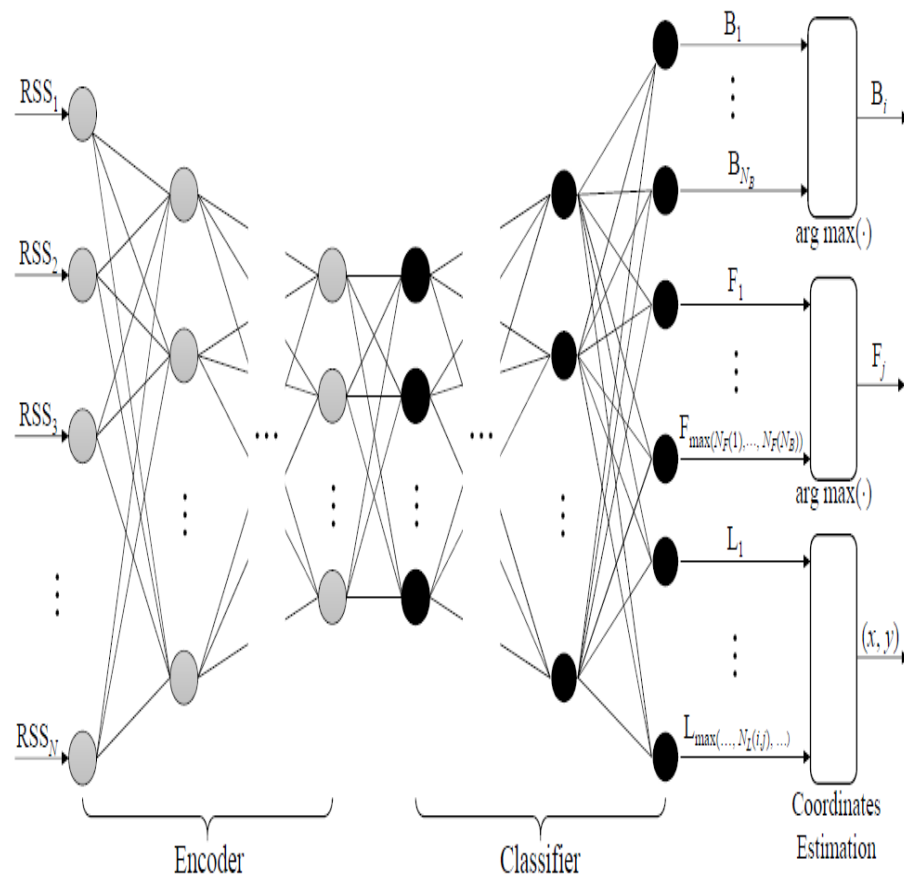


ภาพประกอบ 8 แสดงผลประสิทธิภาพของงานวิจัย Low-effort place recognition with WiFi fingerprints using deep learning

ที่มา : [ออนไลน์] <https://arxiv.org/abs/1611.02049>

ผลการเปรียบเทียบประสิทธิภาพระหว่างโครงสร้างแบบต่าง ๆ ดังรูปที่ 9 จะพบว่าการใช้ SAE จะช่วยให้มีผล Testing Accuracy ดีขึ้น และที่โครงสร้างที่ใช้ SAE (256-128-64) + Classifier (128-128) จะค่า Testing Accuracy สูงที่สุด คือ ประมาณ 0.91 โดยงานนี้จะเป็นการแบ่งแยกอาคารและชั้นเท่านั้น ซึ่งโครงสร้างของ DNN ที่ใช้ จะไม่มีผลลัพธ์เป็นพิกัดตำแหน่ง

งานวิจัยเรื่อง A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting ได้เสนอโครงสร้างของ DNN ซึ่งประกอบด้วย Autoencoder และ Classifier เพื่อใช้ในการแบ่งแยกอาคาร ชั้น และพิกัดตำแหน่ง ( Latitude , Longitude ) โดยมีโครงสร้างดังภาพประกอบที่ 4



ภาพประกอบ 9 แสดงโครงสร้าง DNN ของงานวิจัย A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting

ที่มา : [ออนไลน์] <https://arxiv.org/abs/1712.01990>

โดยมีพารามิเตอร์ที่ใช้ดังตาราง 1

ตาราง 1 แสดงพารามิเตอร์ DNN ของงานวิจัย A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting

DNN Parameter	Value
Ratio of Training Data to Overall Data	0.90
Number of Epochs	20
Batch Size	10
SAE Hidden Layers	256-128-256
SAE Activation	Rectified Linear (ReLU)
SAE Optimizer	ADAM [18]
SAE Loss	Mean Squared Error (MSE)
Classifier Hidden Layers	64-128
Classifier Activation	ReLU
Classifier Optimizer	ADAM
Classifier Loss	Binary Crossentropy
Classifier Dropout Rate	0.20

ที่มา : [ออนไลน์] <https://arxiv.org/abs/1712.01990>

ส่วนของ Classifier ใช้ Loss Function คือ Binary Crossentropy ซึ่งใช้ในการทำ Multi-label Classifications และมีการใช้ Adam Optimizer ซึ่งเป็นวิธีที่มีประสิทธิภาพในการปรับค่า Weight ของ DNN เพื่อให้ค่าตอบลู่เข้าค่าจริงมากที่สุด ในโครงสร้างนี้ใช้อัตรา Dropout ที่ 0.2 เพื่อให้ป้องกันไม่ให้ DNN เกิด Overfitting

ซึ่งงานวิจัยนี้จะนำผลลัพธ์จาก DNN มาผ่านกระบวนการ KNN เพื่อเพิ่มประสิทธิภาพ ซึ่งให้ผลลัพธ์ที่ดีที่สุด เมื่อพารามิเตอร์ของ KNN เป็นดังตารางที่ 1

ตาราง 2 แสดงผลลัพธ์จาก DNN ของงานวิจัย A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting

k	$\sigma$	Building Hit Rate [%]	Floor Hit Rate [%]	Success Rate [%]	Positioning Error [m]	
					Centroid	Weighted Centroid
8	0.2	99.82	91.27	91.18	9.76	9.29

งานวิจัยเรื่อง User Localization in Complex Environments by Multimodal Combination of GPS, Wi-Fi, RFID and Pedometer Technologies ได้เสนอแนวคิดในการใช้เทคโนโลยี Multimodal (GPS, RFID, Pedometer) กับการใช้ Wi-Fi เพียงอย่างเดียวเพื่อทำการเปรียบเทียบประสิทธิภาพของการระบุตำแหน่ง โดยใช้วิธีการเก็บข้อมูลแบบ Grid Points เป็นหลัก ซึ่งผลการทดลองพบว่าค่าความคลาดเคลื่อนของการระบุตำแหน่งโดยเฉลี่ยมีดังนี้

	Average error (m)	Maximal error (m)	Error at reliability of 90% (m)
Multimodal	2.18	6.62	4.07
WiFi only	9.99	23.84	18.44

ภาพประกอบ 10 ตารางการเปรียบเทียบข้อผิดพลาดในการระบุตำแหน่งระหว่าง Multimodal กับ Wi-Fi Only [2]

จากตารางในภาพประกอบที่ 10 จะสังเกตเห็นได้ว่าค่าความคลาดเคลื่อนเฉลี่ยของ Multimodal อยู่ที่ประมาณ 2.18 เมตร ส่วน Wi-Fi only ค่าความคลาดเคลื่อนเฉลี่ยอยู่ที่ประมาณ 9.99 เมตร ดังนั้นจากผลการทดลองดังกล่าวสามารถสรุปได้ว่าการใช้เทคโนโลยีแบบ Multimodal มีค่าความคลาดเคลื่อนต่ำกว่าการใช้ Wi-Fi เพียงอย่างเดียว

งานวิจัยเรื่อง Fusion of Wi-Fi, Smartphone Sensors and Landmarks using The Kalman Filter for Indoor Localization [1] ได้นำเสนอวิธีการรวมเทคโนโลยี Wi-Fi Fingerprinting, Pedestrian Dead Reckoning (PDR) การก้าวเดินของผู้ใช้ และ Identification of Landmarks โดยการนำ Kalman filter ซึ่งปัญหาของ PDR คือในระยะยาวจะมีความแปรปรวนมาก จึงจำเป็นต้องใช้ Landmarks เข้ามาช่วยในการจับการก้าวของผู้ใช้ จะช่วยลดความผิดพลาดได้

งานวิจัยนี้จึงได้นำเสนอการแก้ปัญหา Wi-Fi Fingerprinting โดยการนำ Wi-Fi WPL ที่ใช้ Math model เข้ามาช่วย คือการอ่านค่าสัญญาณแล้วสามารถหาตำแหน่งได้เลย โดยไม่ต้องเก็บข้อมูลให้ยุ่งยาก ซึ่ง Wi-Fi WPL หรือ PDR with Landmarks เพียงอย่างเดียวอาจไม่เพียงพอที่จะทำให้ข้อมูลมีความถูกต้อง แม่นยำ เนื่องจากสัญญาณ Wi-Fi มีการเปลี่ยนแปลง

ตลอดเวลา และ PDR มีความคลาดเคลื่อนตลอดการเดิน ดังนั้นจึงได้นำทั้ง 3 วิธีนี้มาใช้รวมกันบนอุปกรณ์ Smartphone

วิธีที่ใช้	ค่ากลางของความคลาดเคลื่อน (เมตร)
Wi-Fi WPL	3.5189
PDR with landmarks	1.7727
Sensor Fusion with the Kalman filter	<b>0.8492</b>

ภาพประกอบ 11 ตารางผลการทดสอบโดยเปรียบเทียบค่าความคลาดเคลื่อนทั้ง 3 วิธี[1]

จากผลการทดลองพบว่าวิธี Sensor Fusion with the Kalman filter สามารถที่จะช่วยลดค่าความคลาดเคลื่อนได้มากที่สุด โดยที่สามารถช่วยลดค่าความคลาดเคลื่อนของวิธี Wi-Fi WPL ได้ 75.9% และลดค่าความคลาดเคลื่อนวิธี PDR with landmarks ได้ 52.1% ซึ่งค่าความแม่นยำในการระบุตำแหน่ง ค่าเฉลี่ยอยู่ที่ 1 เมตร

งานวิจัยเรื่อง LMDD: Light-weight Magnetic-base Door Detection with Your Smartphone [3] งานวิจัยนี้ได้เสนอการใช้เทคโนโลยี Wi-Fi โดยใช้หลักการ Fingerprinting เพื่อเชื่อมโยงความเข้มของสัญญาณเข้ากับตำแหน่งของผู้ใช้ ซึ่งได้มีการใช้ Landmarks มาช่วยเพิ่มความแม่นยำให้การระบุตำแหน่งมากยิ่งขึ้น

การทำ Landmark ได้ใช้วิธี Infrastructure Free[3] ซึ่งใช้โทรศัพท์มือถือวัดค่าความเข้มของสนามแม่เหล็ก โดยใช้การเปลี่ยนแปลงของสนามแม่เหล็ก ในการที่จะเชื่อมโยงว่ามีการผ่านประตูหรือไม่ ได้ทำการทดลองใน 4 สภาพแวดล้อมที่แตกต่างกัน คือในสำนักงาน ห้องเรียน โรงพยาบาล และบ้าน ผลการทดลองมีดังนี้

Environment	TP	FP	FN	Accuracy
Office	59%	10%	31%	66%
Classroom	61%	19%	20%	75%
Residential house	59%	18%	23%	71%
Hospital	70%	19%	11%	85%

ภาพประกอบ 12 ตารางเปรียบเทียบค่าความถูกต้องเฉลี่ยในสภาพแวดล้อมที่แตกต่างกัน[3]

ผลที่ได้มีค่าความถูกต้องเฉลี่ยรวมอยู่ที่ 74% โดยมีค่าต่ำสุด 66% ที่สำนักงาน และค่าสูงสุด 85% ที่โรงพยาบาล

งานวิจัยเรื่อง Indoor localization with UMTS compared to WLAN [4] ได้นำเสนอการใช้RF-Fingerprint และอุปกรณ์ Small Cells ซึ่งเป็นอีกทางเลือกสำหรับ UMTS ที่จะนำมาใช้วัดสัญญาณภายในอาคาร

วิธีการคือ ทำการเปรียบเทียบ UMTS กับ WLAN ภายในอาคารสำนักงานขนาด 20 x 15 เมตร โดยใช้ 4 Small Cells และ 4 WLAN-APs การกำหนดจุดทำในพื้นที่ขนาด 1 เมตร โดยมีตัวชี้วัดคือ AP และ RSS ซึ่งใช้มาตรฐาน Euclidean distance และมาตรฐาน Kullback-Leibler (KL) ใช้เปรียบเทียบกัน วิธีการระบุตำแหน่งใช้วิธี Weighted K-Nearest-Neighbor algorithm (WKNN)

ผลการทดลองโดยเปรียบเทียบค่าความถูกต้องของ WLAN และ UMTS พบว่าที่ -55 dBm WLAN มีค่าน้อยที่สุดอยู่ที่ 5.5 เมตร และที่ -80 dBm มีค่าน้อยที่สุดอยู่ที่ 5.9 เมตร

Vis. Thres.[dBm]	-85	-80	-75	-65	-60	-55	-50
WLAN [m]	11.8	11.9	9.67	6.3	6.3	5.5	6.2
UMTS [m]	6.19	5.9	13.5	13.9	13.9	-	-

ภาพประกอบ 13 ตารางการเปรียบเทียบค่าความถูกต้องของ WLAN และ UMTS โดยใช้ WKNN เป็นตัวชี้วัด[4]

งานวิจัยเรื่อง Bluetooth Indoor Positioning[5] งานวิจัยนี้ทำการระบุตำแหน่งภายในอาคารโดยอาศัยเทคโนโลยี Bluetooth เพื่อบูรณาการเข้ากับ Global Positioning Module (GPM)

วิธีการคือจะทำการเปรียบเทียบความเข้มของสัญญาณรอบๆบริเวณอุปกรณ์บลูทูธ ซึ่งใช้โทรศัพท์มือถือในการเก็บข้อมูล Latitude, Longitude และ Altitude รวมทั้งการใช้ค่าความเข้มของสัญญาณ

Wi-Fi โดยใช้วิธี k-Nearest Neighbor เปรียบเทียบกับวิธี Naïve Bayes เพื่อทำการทดสอบประสิทธิภาพของการระบุตำแหน่งของผู้ใช้ จากผลการทดลองพบว่าค่าความถูกต้องเฉลี่ยอยู่ที่ 1.5 เมตร

	Accuracy (m)	Precision (m)		
		50%	80%	95%
k-NN	1.62	1.72	3.04	5.16
k-NN regression	1.60	1.42	2.40	3.79
Naïve Bayes	2.13	1.97	3.31	5.72

ภาพประกอบ 14 ตารางการเปรียบเทียบค่าความถูกต้องของ K-NN, K-NN regression และ Naïve Bayes[5]

### บทที่ 3 การพัฒนางานวิจัย

ขั้นตอนเตรียม Dataset ให้พร้อมใช้งาน



ภาพประกอบ 15 แสดงพื้นที่ 3 อาคารของ University Jaume I, Spain

ที่มา : [ออนไลน์] <http://ieeexplore.ieee.org/document/7275492>

เป็นฐานข้อมูลสาธารณะดังรูปที่ 15 ประกอบด้วย 3 อาคาร แต่ละอาคารมี 4-5 ชั้น และมีจำนวนข้อมูล

Training (90%), Validation (10%) : 19,937 (ไฟล์ train.csv)

Testing : 1,111 (ไฟล์ validation.csv)

จำนวน Access point : 520

ประกอบด้วย 3 อาคาร แต่ละอาคารมี 4-5 ชั้น

ฐานข้อมูลนี้เก็บข้อมูลโดยผู้ใช้งานจำนวน 20 คน โดยใช้สมาร์ทโฟนจำนวน 25 เครื่องในการเก็บค่า Fingerprint ซึ่งเป็นความแรงสัญญาณ WiFi จาก Access Point ในแต่ละตำแหน่งภายในแต่ละอาคาร และมีข้อมูลเป็น Sparse Data เนื่องจาก Fingerprint ส่วนใหญ่มีข้อมูล RSSI ของ Access Point ประมาณ 20 ค่า จาก 520 APs ซึ่งสามารถใช้ Stacked Autoencoder เพื่อใช้ในการลดเวลาในการวิเคราะห์ข้อมูลได้

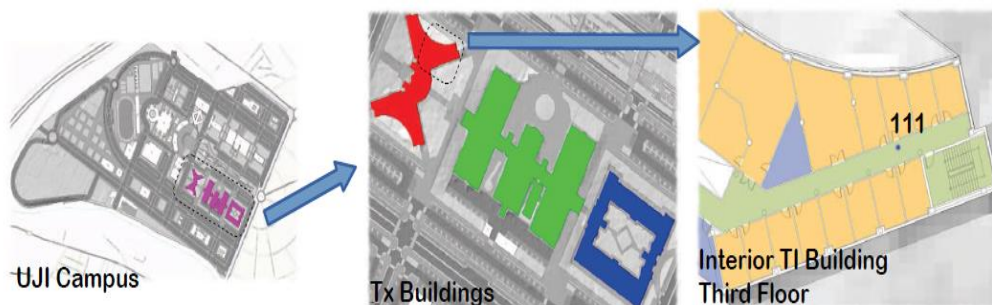
ข้อมูลใน UJIIndoorLoc Dataset ประกอบด้วยค่า RSSI ของ WiFi Access Point ตั้งแต่ WAP1 จนถึง WAP520 ซึ่งเป็นค่าความแรงสัญญาณ WiFi มีหน่วยเป็น dB โดยที่ค่า 100

หมายถึงไม่สามารถรับสัญญาณของ Access Point นั้นได้ ค่า Longitude และ Latitude คือค่าพิกัดตำแหน่งของตำแหน่งที่เก็บข้อมูล ค่า SpacelD คือ ค่ารหัสตำแหน่งซึ่งจะมีการจากแบ่งพื้นที่ภายในชั้นนั้นๆ ค่า Floor คือเลขระบุแต่ละชั้น มีค่าตั้งแต่ 0 ถึง 4 ตามแต่ละอาคาร ค่า Building คือเลขระบุอาคาร ซึ่งอาคารของ University Jaume I ที่ใช้สร้างฐานข้อมูลนี้มี 3 อาคาร ดังนั้นจะใช้ค่า 0 ถึง 2 ในการระบุแต่ละอาคาร ดังตารางที่ 3

ตาราง 3 แสดงตัวอย่างข้อมูลใน UJIIndoorLoc Dataset

WAP1	WAP2	...	WAP52	Longitude	Latitude	SpacelD	Floor	Building
			0					
-94	100	...	100	-7541.26	4864921	106	2	1
100	-91	...	-71	-7524.57	4864934	102	2	1
100	-78	...	100	-7559.78	4864871	108	2	1
-91	100	...	100	-7562.19	4864867	109	2	1

เป็นข้อมูลการเข้าถึง Internet ผ่านตัว WAPs ที่ติดตั้งอยู่ในระบบ Infrastructure โดยเอาข้อมูลการ access เข้าสู่ Internet ของโทรศัพท์มือถือ ณ เวลาที่เคลื่อนที่ผ่าน WAP แต่ละจุดมาใช้ โดย WAPs ถูกติดตั้งไว้ในระบบ Infrastructure ที่ติดตั้งอยู่ในตึกสามตึกใน มหาวิทยาลัย Jaume ใน ประเทศ Spain โดยแต่ละตึกมีจำนวนชั้นอย่างน้อย 4 ชั้น มีพื้นที่อย่างน้อย 110.000 ตารางเมตร. โดย ข้อมูล ที่สามารถใช้ในการ เามาทำ classification เช่น อาคารและชั้น ที่จับสัญญาณได้ , ข้อมูลที่เามาทำ regression เช่น ลองติจูด และ ละติจูด ที่จับสัญญาณได้ ข้อมูล ชุดนี้ถูกการบันทึกเมื่อปี 2013 โดย ใช้ผู้เก็บข้อมูล อย่างน้อย 20 คนและโทรศัพท์มือถือ smartphone ชนิด OS แบบ Android อย่างน้อย 25 เครื่อง. โดยข้อมูล ที่ เก็บได้ มีจำนวน 19937 records (trainingData.csv file) และ 1111 test records (validationData.csv file). โดยในหนึ่ง record ประกอบด้วย 529 attributes เช่น WiFi fingerprint, จุดที่จับสัญญาณได้ และข้อมูล อุปกรณ์ที่ใช้ทดลองและ ตำแหน่ง WAP ( Wireless Access Point หรือตัวกระจายสัญญาณ)

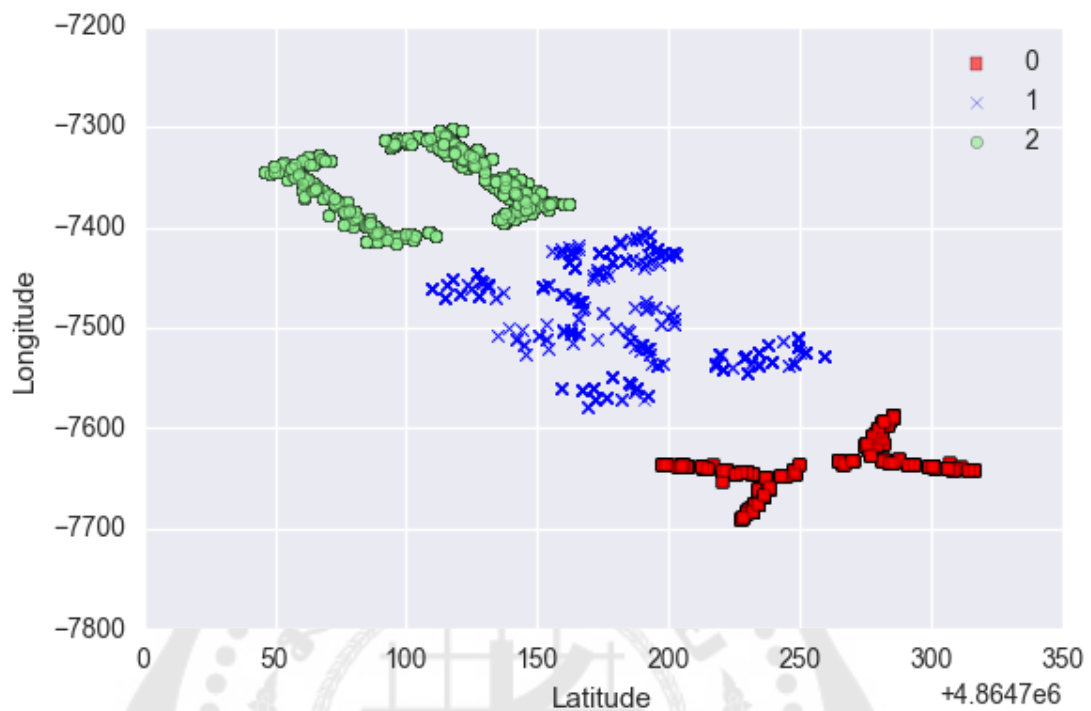


ภาพประกอบ 16 ตารางข้อมูล UJIIndoor

โดย ชื่อคอลัมน์จะเป็นชื่อของตัวกระจายสัญญาณไวไฟจาก ตัวที่หนึ่ง ถึง ตัวที่ห้าร้อย ยี่สิบเก้า ส่วนเลขแถวเป็นตัวแสดงลำดับที่ของการวิ่งของข้อมูลที่ผ่านมาในระบบเครือข่ายสัญญาณไวไฟ โดยจากงานวิจัย[1] จะทำการ Cleaning dataset โดย แบ่งเป็น train และ test โดยใช้ column ที่ 0 ถึง 520 เป็น train และ 521 ถึง 526 เป็น test และ แปลงค่าใน record ที่ access point มีค่า เดิมเป็น 100 ให้เป็น NAN เพราะตั้งค่า 100 ไว้เป็นค่า default กรณีไม่มีโทรศัพท์มือถือเครื่องใด access ผ่าน WAP ตัวนั้น และ เมื่อได้ข้อมูลชุด train และ test แล้วก็ยังไม่สามารถนำข้อมูลไปใช้ใน Machine Learning Model ได้ในทันทีเนื่องจากข้อมูลที่ได้มา ยังมีค่าความ error อยู่สูงจึงต้อง ทำการปรับแต่งข้อมูลตามขั้นตอนทางสถิติ ก่อนโดยทั้ง การกำจัดข้อมูล ที่ไม่ได้อยู่ในลักษณะ Normalization โดยใช้กระบวนการ skewness เพื่อทำการกำจัดข้อมูล ที่ไม่ได้ อยู่ในขอบเขต Normalization ออกไป และ กระบวนการ kurtosis เพื่อกำจัดข้อมูลที่มีค่า ที่สูงหรือ ต่ำโด่ง ผิดปกติออกไป จากนั้นจึงเข้ากระบวนการ box cox อีกครั้งเพื่อกำจัดข้อมูล ส่วนที่มี จำนวน ซ้ำกัน มากๆ ออกไป จากนั้นจึงได้ข้อมูลพร้อมใช้แล้วทำการแบ่งข้อมูลออกเป็น 6 files แล้วจึงนำเข้าสู่ Machine Learning Model เป็นการต่อไป

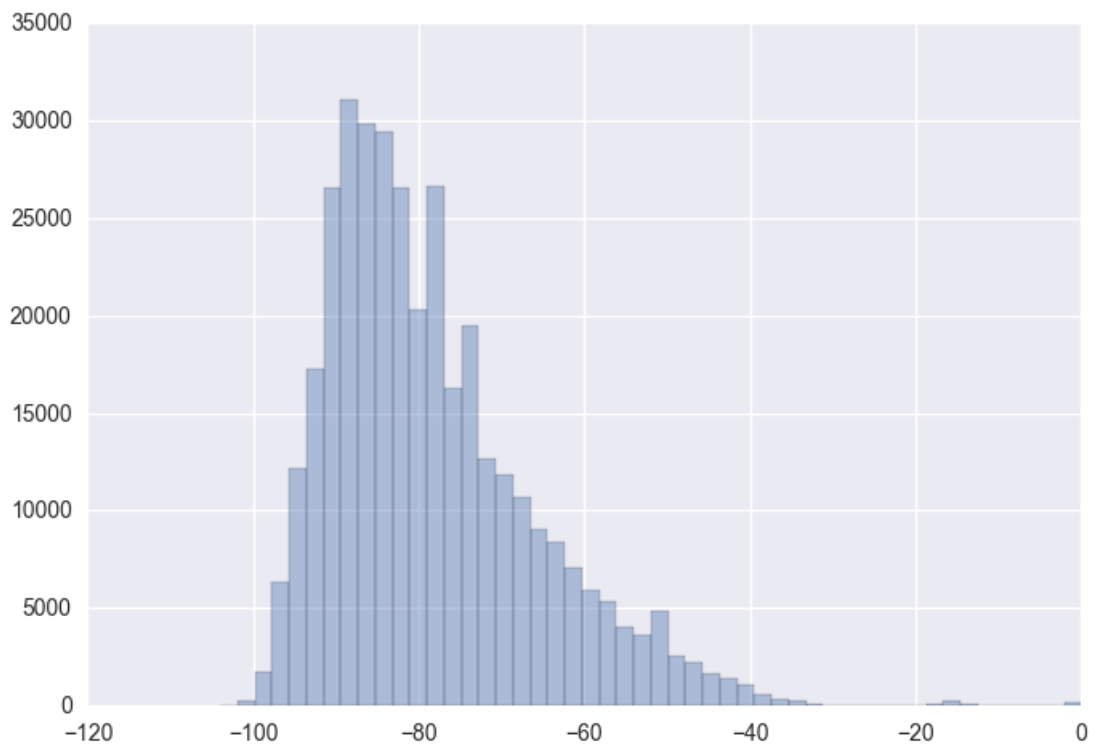
### การทำ Exploratory Data Analysis และ Pre-processing

ในการวิเคราะห์ข้อมูลเบื้องต้นเราได้ทำการวิเคราะห์ข้อมูลความแรงสัญญาณไวไฟเทียบกับตำแหน่งละติจูดและลองติจูดดังแสดงในรูปที่ 18



ภาพประกอบ 17 การกระจายของชุดข้อมูลการ access ของตัว WAPs

จากภาพที่ 18 เป็นการแสดงให้เห็นถึงการกระจายกันของข้อมูลการ access ของ WAPs ในแต่ละอาคาร โดยใช้ Latitude และ Longitude เป็นตัวชี้วัด โดย แสดงเป็นสีเหลี่ยม สีแดงสำหรับอาคารแรก และ กากบาทสีน้ำเงินสำหรับอาคารที่สอง และ วงกลมสีเขียว สำหรับอาคารที่สาม ซึ่งจากรูปจะเห็นได้ว่าความแรงของสัญญาณไวไฟของแต่ละตึกมีการแยกกันตามแต่ละตึกอย่างชัดเจน ไม่มีสัญญาณที่กระจายข้ามอาคารมาได้ ต่อมาเมื่อทำการวิเคราะห์การกระจายตัวของข้อมูล (RSSI) ดังแสดงในรูปที่ 19 พบว่าข้อมูลส่วนใหญ่จะกระจายไปอยู่ทางซ้ายของการกระจายของข้อมูล (right-skewed) มีการกระจายตัวไม่เป็น normal distribution



ภาพประกอบ 18 การกระจายตัวของข้อมูล RSSI

ในส่วนของการหาโมเดลเพื่อวิเคราะห์ตำแหน่งของ smart phone จากค่า WiFi RSSI นั้นเราจะทำการทดสอบประสิทธิภาพของโมเดลจากโมเดลการเรียนรู้ของเครื่องจักรทั้งหมด 3 แบบคือ KNN, Random Forest และ Adaboost โมเดล โดยจะทดสอบประสิทธิภาพของทั้ง 3 โมเดลนี้ในรูปแบบของ Confusion matrix และ accuracy

## บทที่4 ผลการวิจัย

เราจะทำการทดสอบประสิทธิภาพของโมเดล KNN, Random Forest และ Adaboost ในการวิเคราะห์สัญญาณ Wifi เพื่อทำการระบุตำแหน่งของโทรศัพท์มือถือ เนื่องจากความแปรปรวนของสัญญาณ Wifi เราจะระบุตำแหน่งตำแหน่งในรูปแบบของชั้น (floor) ที่พบโทรศัพท์มือถือของแต่ละอาคาร โดยเราจะระบุประสิทธิภาพของโมเดลในรูปแบบของ accuracy และ confusion matrix

ข้อมูลยืนยันเพิ่มเติม เช่น ค่า Accuracy , และ Confusion Matrix ของชุดข้อมูลแยกตามแต่ละ BUILDINGID ซึ่งยืนยันไปในทิศทางเดียวกับงานวิจัย[1]และผลของงานวิจัยนี้ที่ได้กล่าวถึงในย่อหน้าก่อนนี้ โดย BUILDINGID=0 ด้วย model Random Forest มี ค่า Accuracy เท่ากับ 0.7874, และ Confusion Matrix คือ

			PREDICTED LABEL				
			Floor 0	Floor 1	Floor 2	Floor 3	Floor 4
A	L						
C	A	Floor 0	108	102	1	0	N
T	B	Floor 1	0	251	14	1	N
U	E	Floor 2	0	0	290	1	N
A	L	Floor 3	0	0	0	282	N
L	S	Floor4	N	N	N	N	N

ภาพประกอบ 19 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก0และด้วย model K Nearest Neighbors มีค่า Accuracy เท่ากับ 0.9977, และ Confusion Matrix คือ

			PREDICTED LABEL				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	221	0	0	0	N
T	B	FLOOR 1	0	266	0	0	N
U	E	FLOOR 2	0	0	291	0	N
A	L	FLOOR 3	0	0	0	282	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 20 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล KNN ของตึก0 และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788, และ Confusion Matrix คือ

			PREDICTED LABEL				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	211	0	0	0	N
T	B	FLOOR 1	0	0	266	0	N
U	E	FLOOR 2	0	0	291	0	N
A	L	FLOOR 3	0	0	0	282	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 21 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล AdaBoost ของตึก0

โดย BUILDINGID=1 ด้วย model Random Forest มี ค่า Accuracy เท่ากับ 0.7874, และ Confusion Matrix คือ

			PREDICTED LABEL				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	240	15	0	0	N
T	B	FLOOR 1	19	274	7	0	N
U	E	FLOOR 2	0	2	293	0	N
A	L	FLOOR 3	0	40	150	0	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 22 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก1

และด้วย model K Nearest Neighbors มี ค่า Accuracy เท่ากับ 0.9977, และ Confusion Matrix คือ

			PREDICTED				
A	L		FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
C	A	FLOOR 0	225	0	0	0	N
T	B	FLOOR 1	0	299	1	0	N
U	E	FLOOR 2	0	1	293	1	N
A	L	FLOOR 3	0	0	0	190	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 23 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล KNN ของตึก1

และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788, และ Confusion Matrix

คือ

			PREDICTED CLASS				
A	L		FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
C	A	FLOOR 0	255	0	0	0	N
T	B	FLOOR 1	0	300	0	0	N
U	E	FLOOR 2	0	295	0	0	N
A	L	FLOOR 3	0	0	0	190	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 24 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล AdaBoost  
ของตึก1

โดย BUILDINGID=2 ด้วย model Random Forest มี ค่า Accuracy เท่ากับ 0.7874,  
และ Confusion Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	365	7	0	1	0
T	B	FLOOR 1	24	396	0	13	0
U	E	FLOOR 2	0	69	42	201	0
A	L	FLOOR 3	0	0	0	558	0
L	S	FLOOR 4	0	0	0	165	58

ภาพประกอบ 25 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก2

และด้วย model K Nearest Neighbors มีค่า Accuracy เท่ากับ 0.9977, และ Confusion Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	373	0	0	0	0
T	B	FLOOR 1	1	432	0	0	0
U	E	FLOOR 2	0	3	308	1	0
A	L	FLOOR 3	0	0	3	558	0
L	S	FLOOR 4	1	0	0	0	222

ภาพประกอบ 26 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล KNN ของตึก2

และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788, และ Confusion Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	373	0	0	0	0
T	B	FLOOR 1	0	433	0	0	0
U	E	FLOOR 2	0	312	0	0	0
A	L	FLOOR 3	0	0	0	558	0
L	S	FLOOR 4	0	0	0	0	223

ภาพประกอบ 27 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล AdaBoost ของตึก2

จากการวิเคราะห์ผลการทดลองจะเห็นได้ว่าโมเดล KNN มีประสิทธิภาพดีที่สุดในการทำนายตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi โดยมีประสิทธิภาพดีที่สุดในการทำนายตำแหน่งของอาคาร 0, 1 และ 2 ตามลำดับ โดยผลจากโมเดล Random Forest และ Adaboost ก็มีแนวโน้มคล้ายๆกันซึ่งสาเหตุที่น่าจะเป็นไปได้คือสัญญาณ WiFi ที่อาคาร 2 มีความไม่แน่นอนสูงทำให้ผลลัพธ์ที่ได้มีความคลาดเคลื่อนเยอะ ดังนั้นอาจจะต้องมีการประมวลผลสัญญาณเพื่อลดความไม่แน่นอนของสัญญาณก่อนจะนำเข้ามาใส่ในโมเดลการเรียนรู้ของเครื่องจักรต่อไป



## บทที่ 5

### สรุปและแนวทางการพัฒนาที่จะทำเพิ่มเติม

#### สรุป

งานวิจัยนี้เสนอการใช้เทคนิคการเรียนรู้ของเครื่องจักรในการระบุตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi ภายในอาคาร จากการทดลองพบว่าโมเดล KNN ให้ความแม่นยำมากที่สุดในการทำนาย โดยมีความถูกต้องสูงถึง 97% ในการระบุตำแหน่งในรูปของชั้นในอาคารที่โทรศัพท์มือถืออยู่ อย่างไรก็ตามการระบุพิกัดของโทรศัพท์มือถือในรูปของละติจูดและลองจิจูดยังมีความคลาดเคลื่อนสูงอยู่ซึ่งมีสาเหตุหลักมาจากความแปรปรวนของสัญญาณ WiFi ภายในอาคารซึ่งแนวทางแก้ไขที่เป็นไปได้คือการกรองสัญญาณ WiFi ก่อนนำเข้าโมเดลการเรียนรู้ของเครื่องจักร หรือ การใช้สัญญาณแบบอื่น เช่น สัญญาณ Beacon มาช่วยในการระบุตำแหน่งให้มีความแม่นยำมากขึ้นต่อไป

#### การพัฒนาที่น่าจะทำเพิ่ม

ศึกษาแนวทางการประมวลผลสัญญาณเพื่อลดผลของสัญญาณรบกวนและลดความแปรปรวนของสัญญาณก่อนจะนำเข้าสู่โมเดลการเรียนรู้ของเครื่องจักรต่อไป

## บรรณานุกรม

1. Feng, Y., et al. *Improved AdaBoost-based fingerprint algorithm for WiFi indoor localization*. in *2014 IEEE 7th joint international information technology and artificial intelligence conference*. 2014. IEEE.
2. Rusli, M.E., et al. *An improved indoor positioning algorithm based on rssi-trilateration technique for internet of things (iot)*. in *2016 International Conference on Computer and Communication Engineering (ICCCE)*. 2016. IEEE.
3. Torres-Sospedra, J., et al. *UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems*. in *2014 international conference on indoor positioning and indoor navigation (IPIN)*. 2014. IEEE.
4. Xu, X., Y. Tang, and S. Li. *Indoor localization based on hybrid Wi-Fi hotspots*. in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 2017. IEEE.
5. Improved AdaBoost-based fingerprint algorithm for WiFi indoor localization , 2014 IEEE 7th joint international information technology and artificial intelligence conference
6. An improved indoor positioning algorithm based on rssi-trilateration technique for internet of things , 2016 International Conference on Computer and Communication Engineering (ICCCE)
7. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems , 2014 international conference on indoor positioning and indoor navigation (IPIN)
8. Indoor localization based on hybrid Wi-Fi hotspots , 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)
9. Sharan Naribole , A Machine Learning Approach to Wi-Fi Fingerprint based

Localization,2014

<https://github.com/sharan-Torres-Sospedra>, 2014

[#1}naribole/wlan\\_localization/blob/master/UJIIndoorLoc-machine-learning.ipynb](#)

- 10 Yu Feng , Jiang Minghua , Liang Jing , Hu Ming , Peng Tao , Hu Xinrong ;  
,[Improved AdaBoost-based fingerprint algorithm for WiFi indoor localization](#) ; ,  
2015





## ประวัติผู้เขียน

ชื่อ-สกุล	ปฏิพล ศรีชุ่มสิน
วัน เดือน ปี เกิด	26 กุมภาพันธ์ 2522
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	พ.ศ. 2563 วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ จาก มหาวิทยาลัยศรีนครินทรวิโรฒ
ที่อยู่ปัจจุบัน	เลขที่ 10 , ซอย รามคำแหง118 แยก19 , เขต สะพานสูง , แขวง สะพาน สูง , กรุงเทพฯ 10240

