Nama / Name : Nuraisa Novia Hidayati

Nrp / Student Number : 05111950010019

Judul Paper / Paper Title :

1. Estimating Room Occupancy in a Smart Campus using WiFi Soft Sensors

Pengarang / Authors:

1. Iresha Pasquel Mohottige

Penerbit Publikasi dan Tahun (Publisher and Year) :

1. 2018 / IEEE 43rd Conference on Local Computer Network

# 

# Summary / Abstrak

* Estimating Room Occupancy in a Smart Campus using WiFi Soft Sensor

Terdapat lonjakan jumlah mahasiswa pada kampus UNSW, membuat pengelola ruangan membutuhkan data untuk melakukan optimalisasi hunian pada setiap ruangan yang ada. Pengukuran kepadatan hunian pada sebuah ruangan biasanya menggunakan instrumentasi sensor perangkat keras seperti sensor lingkungan (CO2, suhu, kelembaban) atau kamera video untuk melakukan proses terhadap keberadaan penghuni, yang keduanya memiliki kelemahan: pada kasus pertama umumnya memiliki akurasi yang rendah, sedangkan pada kasus kedua membutuhkan biaya pengaturan dan pemeliharaan yang tinggi

Sementara itu, pada penelitian kali ini diambil pendekatan yang berbeda dengan menggunakan data konektivitas Wifi. Konektivitas Wifi sendiri bukanlah alat prediksi yang baik untuk tingkat hunian dalam sebuah ruangan. Hal ini dikarenakan data konektivitas dapat tercampur dengan ruangan lain yang berdekatan, jalan atau titik berkumpul mahasiswa di luar ruangan terdekat, dan adanya penyeimbang beban jaringan.

Berdasarkan data koneksi AP pada setiap ruangan, dilakukan analisa mengenai karakteristik data yang berasal dari koneksi penghuni (occupant) atau berasal dari koneksi orang yang berada diluar ruangan yang berdekatan (bystander). Data kemudian di klasifikasikan berdasarkan kelas occupant dan bystander. Data awal yang telah terkumpul kemudian menjadi bahan training pada beberapa algoritma machine learning, untuk kemudian dilakukan testing pada data selanjutnya.

Analisa Regresi kemudian diterapkan pada kelompok data yang di prediksi sebagai occupant dengan data occupant sebenarnya untuk menghitung estimasi tingkat hunian ruangan. Evaluasi terhadap teknik ini dilakukan pada berbagai ruang kuliah dengan berbagai ukuran di kampus UNSW, dan menunjukkan bahwa keakuratannya mendekati teknik yang menggunakan sensor perangkat keras.

# Latar Belakang (Review Paper)

## Permasalahan Penelitian

* Estimating Room Occupancy in a Smart Campus using WiFi Soft Sensor

Mahasiswa yang menghadiri kelas akan bervariasi jumlahnya, tergantung pada waktu, keterlibatan dosen, dan ketersediaan konten online. Bagaimanapun, alokasi ruangan kelas di atur berdasarkan pendaftaran. Pada kenyataannya cukup banyak kasus yang membuktikan bahwa banyaknya mahasiswa yang hadir dalam kelas berbeda dengan jumlah mahasiswa yang mendaftar pada kelas tersebut. Hal ini seharusnya menjadi perhatian pengelola untuk memaksimalkan alokasi ruangan tidak lagi berdasarkan pada data saat pendaftaran, tetapi berdasarkan level kehadiran mahasiswa sesungguhnya. Perangkat keras berupa sensor yang banyak di aplikasikan untuk mengetahui tingkat hunian/kepadatan pada suatu lokasi membutuhkan biaya dan usaha yang tidak kecil untuk pengembangan dan pemeliharaan. Dikarenakan pertumbuhan kepemilikan perangkat mobile semakin pesat maka di gagaslah penggunaan konektivitas Wifi sebagai sensor tingkat hunian.

Meskipun Infrastruktur WiFi telah dipergunakan sebagai gambaran kasar atas kondisi spasial seperti lantai pada suatu bangunan dan tingkat huni pada ruangan kecil. Akan tetapi, dalam penggunaannya sebagai alat prediksi masih belum berhasil ketika diterapkan pada ruangan besar seperti teater universitas. Sinyal WiFi yang menyebar ke ruangan lain merupakan tantangan utama dalam penerapannya sebagai alat prediksi tingkat huni. Pengguna di luar ruangan yang tidak mengikuti perkuliahan dapat melakukan akses ke AP pada ruangan tersebut saat jaraknya berdekatan, selain itu mahasiswa yang berada di dalam ruangan juga dapat melakukan koneksi dengan lebih dari satu perangkat. Dua kasus tersebut dapat menimbulkan estimasi berlebihan terhadap tingkat huni. Kemudian mahasiswa yang berada dalam ruangan tetapi tidak memiliki perangkat yang terhubung akan menimbulkan kekurangan pada estimasi tingkat huni.

## Kontribusi Utama

* Estimating Room Occupancy in a Smart Campus using WiFi Soft Sensor

Kontribusi pertama adalah usulan data set yang kaya akan fitur dari konektivitas WiFi membedakan antara penghuni dan bukan penghuni. Kontribusi kedua adalah klasifikasi dan analisis regresi berdasarkan pendekatan supervised learning untuk melakukan estimasi tingkat huni pada suatu ruangan menggunakan fitur yang diusulkan.

## Penelitian yang berkaitan

* Estimating Room Occupancy in a Smart Campus using WiFi Soft Sensor

1. Menggunakan Machine Learning seperti SVM, NN, dan Hidden Markov Models(HMM) sebagai pengolah data yang berasal dari jaringan sensor yang terdiri dari sensor CO2 dan sensor perubahan lingkungan sekitar. HMM memberikan hasil yang paling mendekati untuk memprediksi jumlah penghuni dengan akurasi 73%, tetapi hanya pada ruangan kecil dengan kapasitas kurang dari 10 orang [2].
2. Sensor Infrared tunggal yang pasif dengan MAE satu, tetapi hanya pada ruangan dengan kapasitas 14 orang atau kurang [3].
3. Image processing pada kamera, MAE 1.15. Algoritma untuk pengembangannya membutuhkan sumber daya komputasi yang berat [4].
4. Sensor kamera dan sensor termal yang dikombinasikan dengan SVR (support vector Regression). Berhasil melakukan prediksi dengan akurat 7 orang penghuni dari rentang 0-150 pada ruangan yang cukup luas. Akan tetapi, hanya dapat berfungsi dengan baik ketika tidak banyak didapati gerakan [5].
5. Aplikasi mobile untuk mengumpulkan RSSI (Received Signal Strength Indication) data beacons yang di transmisikan dari Apples iBeacons [6]. Kemudian penelitian selanjutnya memodifikasi protocol iBeacon. Keduanya berhasil melakukan estimasi dengan akurasi mendekati 100% tetapi membutuhkan kerjasama dari penghuni ruangan [7].
6. Menggunakan beberapa perangkat WiFi (Raspberry Pi) dalam sebuah ruangan untuk mengumpulkan RSSI dari jaringan WiFi. Menghubungkan tingkat hunian dalam sebuah ruangan dengan perubahan pada propagasi sinyal antara APs dan perangkat menggunakan Linear Regression (LR) dan SVR untuk mencapai MAE 0.471 pada ruangan berkapasitas maksimum 8 orang [8].
7. WiFi sidik jari, walau bagaimanapun membutuhkan kerjasama penghuni/partisipan. Penelitian lainnya menggunakan pengawasan data MAC dan IP address pada Router dan WiFi APs. Cukup berhasil dalam memprediksi hunian dalam satu bangunan, tetapi gagal ketika harus merinci pada setiap laintai nya. Hal ini dikarenakan tumpang tindih area cakupan AP dan konektivitas nirkabel yang tidak konsisten pada setiap perangkat. Tingkat akurasi kemudian di tingkatkan pada penelitian yang lain menggunakan alamat MAC WiFi dan lokasi AP dari Log WiFi. Akurasi berhasil ditingkatkan hingga mencapai 86% pada penentuan tingkat huni sebuah ruangan dalam bangunan komersial [11].
8. Dua penelitian sebelumnya yang paling mendekati penelitian ini telah melakukan ekstraksi fitur dari aktivitas WiFi seperti kualitas sinyal , jumlah koneksi, dll. Penelitian pertama mengambil data dalam rentang waktu yang berbeda dalam satu hari untuk melihat tingkat huni menggunakan metode klasifikasi , yaitu logistic regression dan Linear Discriminant Analysis (LDA) . LDA menunjukkan akurasi yang tinggi yaitu 92% dalam menentukan ruangan kosong atau tidak kosong. Sementara itu, penyaringan data ambigu seperti pengguna di luar ruangan dan melakukan koneksi dilihat dari threshold RSSI. Akan tetapi tidak mengukur efektivitas RSSI atau melihat kemungkinan pendekatan berbeda [13]. Kemudian penelitian kedua melihat adanya tren untuk menggunakan infrastruktur yang merekam aktivitas seperti WiFi menggantikan sensor yang lebih rumit dengan biaya mahal. Penelitian kedua juga berhadapan dengan tantangan untuk melakukan akurasi prediksi pada tempat dengan kapasitas besar [14].

Penelitian ini kemudian menyimpulkan, bahwa merekalah yang pertama kali melakukan ekstraksi metadata aktivitas WiFi dan menggabungkannya dengan metode machine learning untuk memprediksi tingkat hunian dalam ruangan berkapasitas besar yang terletak di dalam kampus.

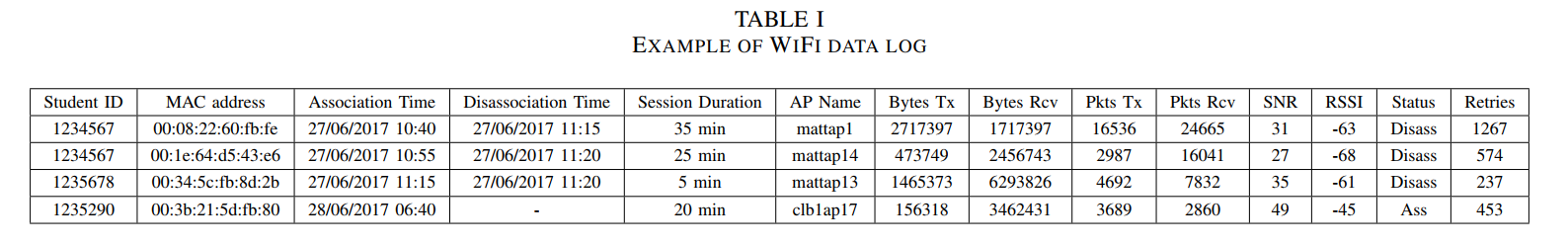
Kedua penelitian merupakan penyempurnaan terhadap penelitian sebelumnya dengan mengubah objek penelitian yang mempengaruhi tingkat hunian. Proses yang dilalui hampir sama, yaitu mengumpulkan data terkait objek yang diteliti kemudian ditambahkan proses machine learning untuk klasifikasi atau regressi untuk prediksi. Dapat dikatakan kontribusi keduanya berada pada level medium, dikarenakan memberikan ide objek penelitian baru yang dapat menangani permasalahan yang timbul akan penelitian-penelitian sebelumnya.

# Bagaimana Metodologi Penelitian Dilakukan / Research Metodology

## Perumusan dan pemodelan permasalahan secara rinci

* Estimating Room Occupancy in a Smart Campus using WiFi Soft Sensor

Data koneksi WiFi diperoleh melalui 41 AP yang terpasang pada 7 ruangan kelas. Contoh format data yang di dapat terlihat pada tabel di bawah ini :



Penjelasan tiap komponen sebagai berikut :

1. Student ID, merupakan identitas pengguna yang terkoneksi. Identitas ini nantinya digunakan untuk melihat beberapa pengguna yang menggunakan lebih dari satu perangkat.
2. MAC address, merupakan identifikasi perangkat yang terhubung. Seperti pada contoh data, terdapat satu pengguna yang menggunakan dua perangkat berbeda. Hal ini diketahui dari Student ID yang sama, tetapi MAC address yang berbeda.
3. AP Name, merupakan penamaan perangkat WiFi. Pada lingkungan UNSW setiap AP diberi nama sesuai lokasi penempatannya.
4. RSSI, merupakan rata-rata kekuatan penerimaan sinyal.
5. SNR, merupakan sinyal rata-rata dengan rasio noise.
6. Status, menandai kondisi perangkat terhubung (Ass) atau sudah tidak terhubung (Disass). Karena laporan data WiFi di buat pada jam 7 pagi, maka mayoritas statusnya adalah tidak terhubung.
7. Retries mengindikasikan berapa kali data frames telah terkirim ke receiver sampai AP menerima ACK (acknowledgement) selama sesi.

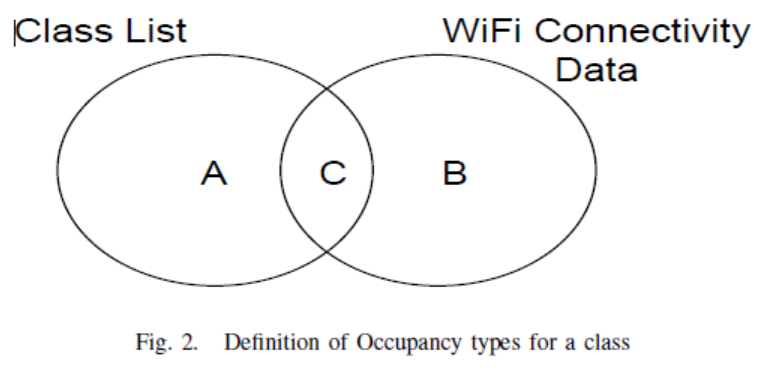
Karena data WiFi yang diperoleh mengandung data-data pribadi seperti username pengguna (Student ID) dan identitas perangkat pengguna, maka pada penelitian ini menerapkan ethical clearance (UNSW Human Research Ethics Advisory Panel Approval number HC17140) untuk menggunakan data-data tersebut. Untuk itulah, dicantumkan ketentuan bahwa semua koneksi WiFi di pantau, dengan menyetujui ketentuan tersebut , pengguna telah memberikan izin untuk menggunakan data-data pribadinya yang tercantum selama koneksi WiFi terjadi.

Adanya identitas pengguna yang unik memudahkan untuk mengenali pengguna yang menggunakan perangkat lebih dari satu, sehingga tidak terjadi perhitungan ganda [15]. Dasar perhitungan awal untuk menentukan klasifikasi menggunakan data mahasiswa yang terdaftar, sementara untuk menentukan jumlah penghuni ruangan sebenarnya dilakukan perhitungan manual melalui jasa sukarelawan.

Data dikumpulkan dari 40 jadwal kelas, kemudian dibandingkan jumlah mahasiswa yang mendaftar dengan mahasiswa yang benar-benar hadir di kelas. Pada sebagian besar jadwal kelas jumlah mahasiswa yang hadir lebih sedikit daripada jumlah mahasiswa yang mendaftar. Hal ini membuka kesempatan optimalisasi ruang kelas dengan melihat pola kehadiran.

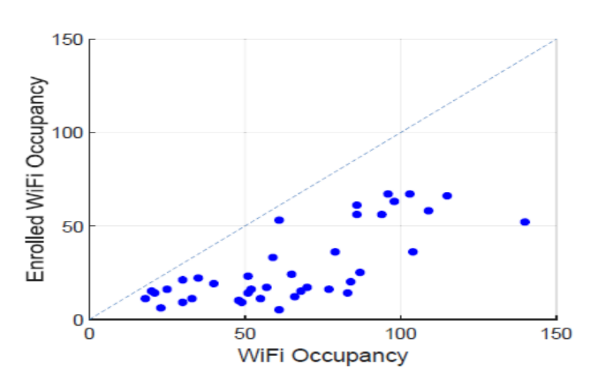
## Algoritma dan implementasi

Data pengguna yang terkoneksi dengan WiFi terdiri dari dua tipe pengguna. Pertama adalah pengguna yang benar-benar berada dalam ruangan, disebut occupant. Kedua adalah pengguna yang berada di luar ruangan, namun memiliki jarak yang dekat dengan AP WiFi ruangan tersebut, disebut bystander. Dengan adanya daftar mahasiswa pada setiap kelas, bystander di filter dengan melakukan asumsi bahwa pengguna yang tercantum pada daftar mahasiswa dan muncul pada data koneksi WiFi adalah penghuni ruangan. Irisan dari mahasiswa yang terdaftar dengan pengguna WiFi adalah penghuni ruangan.

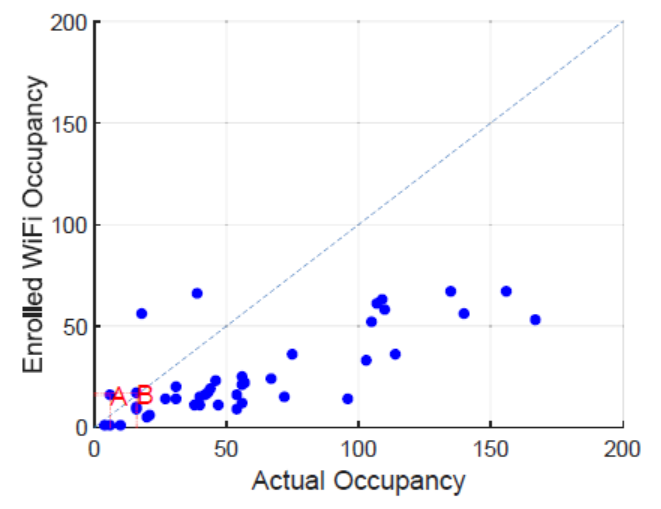


Berdasarkan data yang dikumpulkan juga ditemukan bahwa :

1. Jumlah mahasiswa terdaftar yang terkoneksi WiFi selalu lebih sedikit daripada jumlah pengguna koneksi WiFi pada AP kelas tersebut pada jam kelas diadakan.



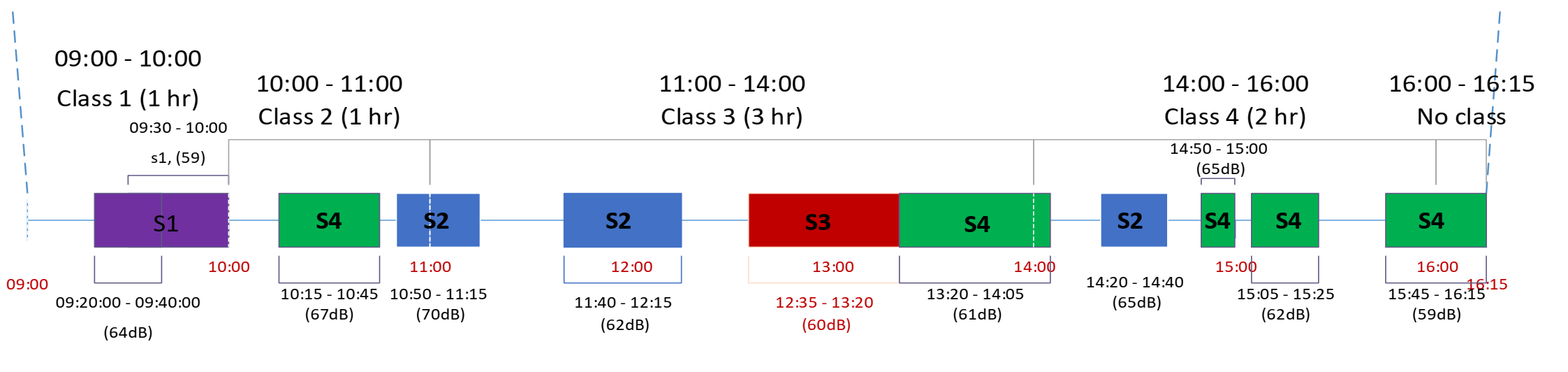
1. Jumlah mahasiswa terdaftar yang terkoneksi WiFi lebih sedikit daripada jumlah penghuni ruangan sebenarnya. Hal ini dikarenakan mahasiswa yang hadir di kelas tidak memiliki perangkat yang terhubung dengan WiFi kelas.
2. Jumlah mahasiswa terdaftar yang terkoneksi WiFi lebih banyak daripada jumlah penghuni ruangan sebenarnya. Hal ini jarang terjadi dan dianggap anomaly karena terjadi ketika mahasiswa yang terdaftar tidak menghadiri kelas, namun berada di sekitar kelas dan melakukan koneksi WiFi melalui perangkatnya.

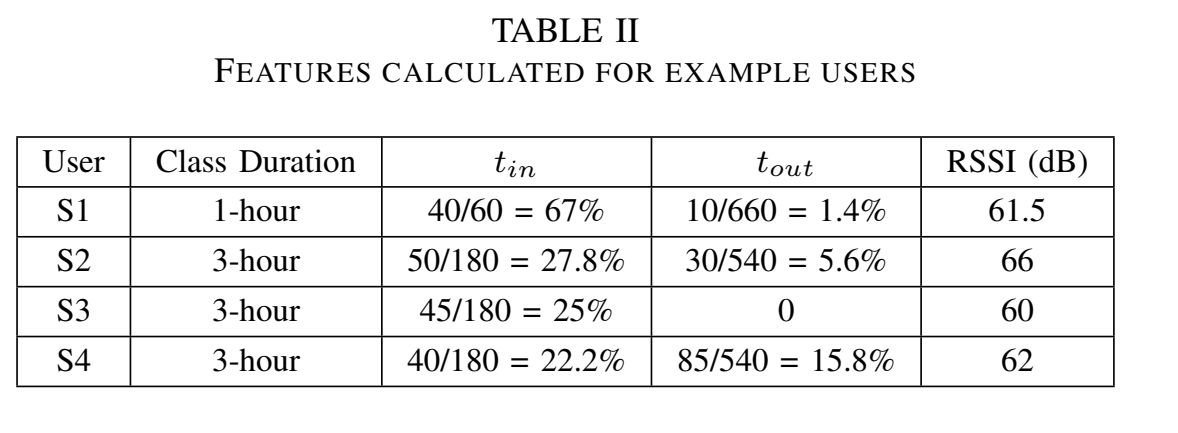


Data penghuni ruangan sebenarnya memiliki hubungan yang lebih baik dengan data mahasiswa terdaftar yang terkoneksi WiFi daripada data pengguna WiFi secara keseluruhan. Hal ini menunjukkan bahwa bystander dapat di filter menggunakan daftar mahasiswa pada kelas tersebut. Akan tetapi, permasalahan muncul ketika ruangan dipakai oleh kegiatan tidak terjadwal seperti rapat dan seminar yang tidak memiliki daftar mahasiswa.

Oleh karena itulah, pada penelitian kali ini filtering bystander di fokuskan pada fitur yang telah di ekstraksi dari data koneksi WiFi sebagai berikut :

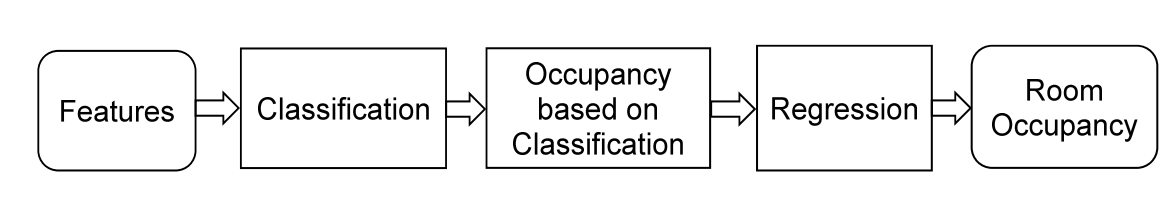
1. RSSI, bystander di prediksi akan menerima sinyal yang lebih lemah daripada occupant.
2. Arrival Delay, waktu koneksi WiFi bystander akan di prediksi akan sangat telat dibandingkan dengan occupant. Patokan waktu awal koneksi adalah disaat jam kelas dimulai.
3. Number of Session, occupant di prediksi akan terkoneksi dengan WiFi dalam waktu lama. Sehingga akan memiliki multiple asosiasi dikarenakan ketidak konsistenan koneksi WiFi pada perangkat mobile.
4. Number of device, bystander biasanya hanya melewati kelas atau berada di dekat kelas tersebut. Sehingga mereka cenderung hanya melakukan koneksi dengan satu perangkat sementara occupant akan melakukan koneksi dengan banyak perangkat (tablet, laptop, hand phone).
5. Persentasi “in time”, persentasi lama koneksi WiFi pada jam kelas. Bystander di prediksi akan berada dalam koneksi WiFi pada waktu yang lebih pendek di saat jam kelas dibandingkan dengan occupant
6. Persentasi “out time”, persentasi lama koneksi WiFi diluar jam kelas. Bystander akan melakukan koneksi WiFi diluar jam kelas lebih sering dibandingkan dengan occupant.





Berdasarkan fitur-fitur diatas , maka diperlukan adanya perbandingan nilai yang tercantum seberapa besar korelasinya degan perbedaan perilaku bystander dan occupant. Berdasarkan analisa yang ada, Occupant akan memiliki nilai “in time” yang tinggi (27.3% - 67.9%), “out time” yang rendah (3% - 25.1%), dan arrival delay yang rendah pula. Kemudian occupant menggunakan perangkat lebih dari satu, serta memiliki banyak sesi dalam range waktu jam kelas. Sementara itu, RSSI tidak memberikan perbedaan yang signifikan terhadap distribusi occupant dan bystander.

Sehingga metodologi yang akan diterapkan ditunjukkan pada gambar di bawah ini :



Pertama adalah mengenali fitur yang ada dari koneksi WiFi selama jam kelas. Kemudian membuat klasifikasi bystander dan occupant. Setelah itu diterapkan beberapa algoritma machine learning untuk membedakan occupant dan bystander. Kemudian mengambil data occupant untuk kemudian diterapkan kembali algoritma regressi untuk melihat kemungkinan tingkat hunian. Terakhir diketahuilah kemungkinan penghuni yang ada pada sebuah ruangan kelas.

# Bagaimana Ujicoba Dilakukan

## Klasifikasi Pengguna WiFi

Terdapat 10000 pengguna WiFi dari berbagai kelas. Metode klasifikasi yang digunakan adalah logistic regression, SVM, dan LDA, ketiganya telah dipakai secara luas untuk memproses binary data. Ketiga metode ini akan dibandingkan performansinya untuk membedakan occupant dan bystander. Perlu diperhatikan hal klasifikasi ini bertujuan untuk melakukan filtering data nantinya terhadap bystander.

Untuk setiap pengguna WiFi, fitur data yang akan dipergunakan adalah sebagai berikut :

1. Persentasi “in time”, persentasi waktu terhubung koneksi WiFi pada jam kelas dibandingkan dengan jam kelas keseluruhan.
2. Persentasi “out time”, persentasi waktu terhubung koneksi WiFi diluar jam kelas dibandingkan dengan jam tanpa kelas secara keseluruhan.
3. Arrival delay, selisih waktu koneksi WiFi dengan saat jam kelas dimulai.
4. Number of sessions, sesi koneksi WiFi.
5. Number of device, jumlah perangkat terkoneksi.

## Regression Analysis

Klasifikasi data untuk menentukan occupant belum terlalu sempurna, dikarenakan adanya beberapa situasi seperti pengguna yang berada di dalam kelas tidak memiliki perangkat atau menggunakan koneksi internet pribadi. Untuk itulah analisis regressi diberlakukan.

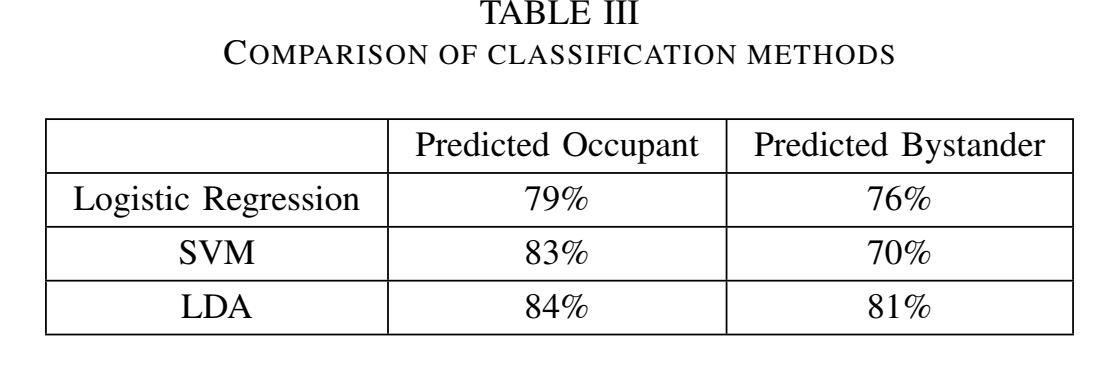
Selama 12 minggu, 790 jadwal kelas pada mata kuliah berbeda-beda dan 7 ruangan kelas pada UNSW. 46% kelas berdurasi 1 jam, 43% kelas berdurasi 2 jam, 8% kelas berdurasi 3 jam, 2% berdurasi 1,5 jam, dan 1% berdurasi 4 jam. Dalam jangka waktu tersebut semua kelas di alokasikan untuk proses belajar mengajar, selain itu terdapat pula kegiatan ujian. Pada kegiatan ujian akan hampir tidak ada aktivitas koneksi WiFi. Sehingga data selama ujian berlangsung tidak dipergunakan. Pada setiap kelas, jumlah pengguna yang masuk pada klasifikasi occupant akan menjadi independent variabel terhadap jumlah penghuni sesungguhnya sebagai dependent variabelnya.

Metode yang digunakan pada analisis regressi ada 2, yaitu :

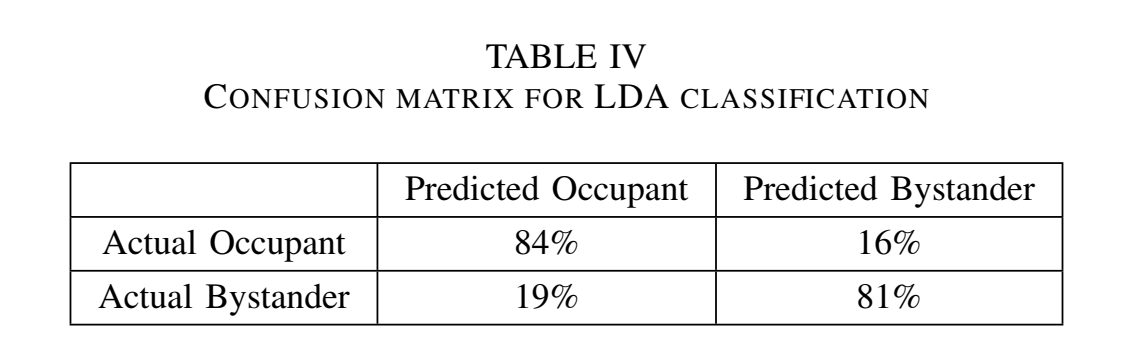
* Linear Regression, mengetahui hubungan antara satu atau lebih independent variable terhadap dependent variable secara linear.
* Support Vector Regression, adalah bentuk regressi lain yang sering digunakan dan merupakan perbaikan dari Linear Regression.

## Evaluasi

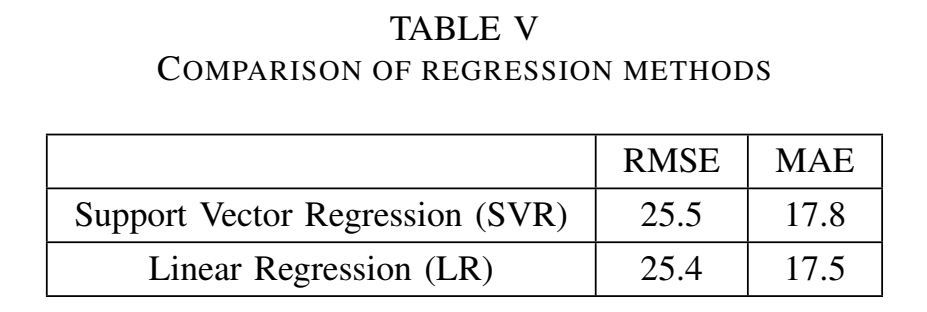
Untuk melakukan evaluasi terhadap metode klasifikasi yang digunakan , terdapat 1500 pengguna WiFi sebagai data uji nya (test set). Prediksi diberlakukan terhadap dua kelas, yaitu occupant dan bystander. Diantara metode klasifikasi tersebut LDA menunjukkan performansi yang paling baik. Untuk melakukan klasifikasi terhadap actual occupant , model tersebut berhasil mencapai akurasi 84%, sementara itu klasifikasi terhadap actual bystander jatuh hingga 81%. Perbandingan nilai akurasi pada masing-masing metode dijelaskan pada tabel berikut :



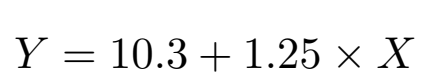
Tingkat kesalahan prediksi dijelaskan pada confusion matrix berikut :



Jumlah pengguna yang berhasil di klasifikasikan ke dalam occupant kemudian akan dimasukkan sebagai input pada analisis regressi. Data ini kemudian akan disandingkan dengan data jumlah penghuni sebenarnya untuk melihat hubungan antara keduanya. Sebagai alat ukurnya, digunakan kalkulasi Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE). Berdasarkan kedua metode tersebut LR menunjukkan performansi lebih baik dengan nilai seperti pada tabel dibawah ini :

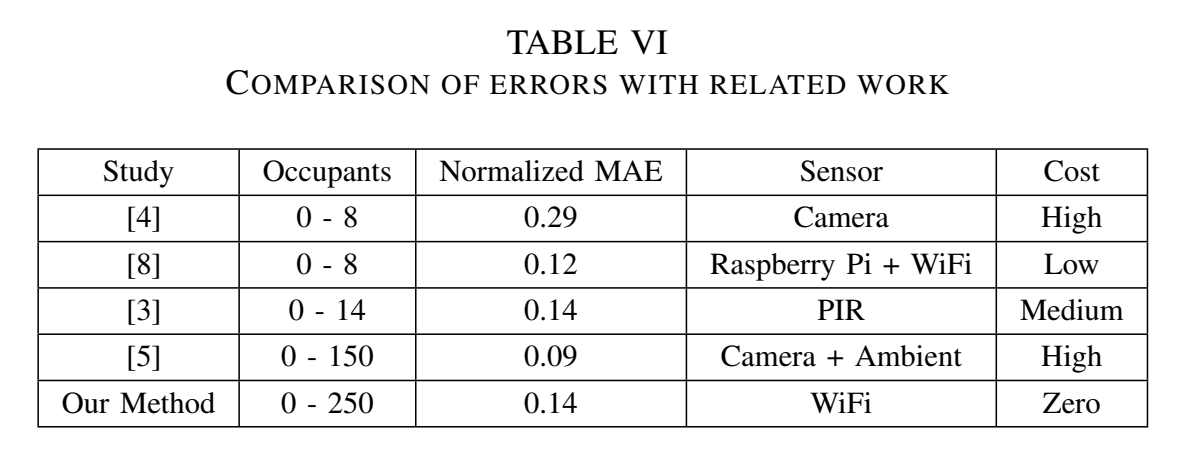


Formula yang didapat oleh Linear Regression terhadap data latihan (training dataset) adalah sebagai berikut :



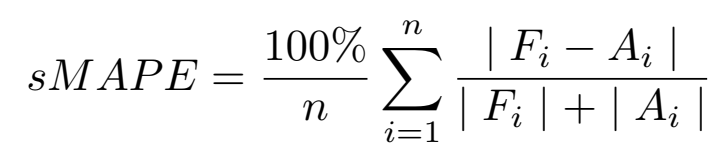
Dimana Y adalah jumlah penghuni sebenarnya dan X adalah jumlah penghuni yang di prediksi berdasarkan klasifikasi LDA. Hasil LR memperlihatkan bahwa X harus mengalami peningkatan sedemikian rupa untuk mendekati nilai Y. Seperti yang telah diprediksi sebelumnya bahwa jumlah penghuni sebenarnya cenderung lebih besar dari jumlah pengguna yang terkoneksi WiFi dan terdaftar pada jadwal kelas di ruangan tersebut.

Performansi metode tersebut kemudian di bandingkan dengan beberapa penelitian yang sejenis. Tolak ukurnya adalah tingkat errornya dengan kalkulasi mean absolute error (MAE). Beberapa tolak ukur lain juga disajikan seperti biaya yang harus dikeluarkan untuk melakukannya sebagai berikut :

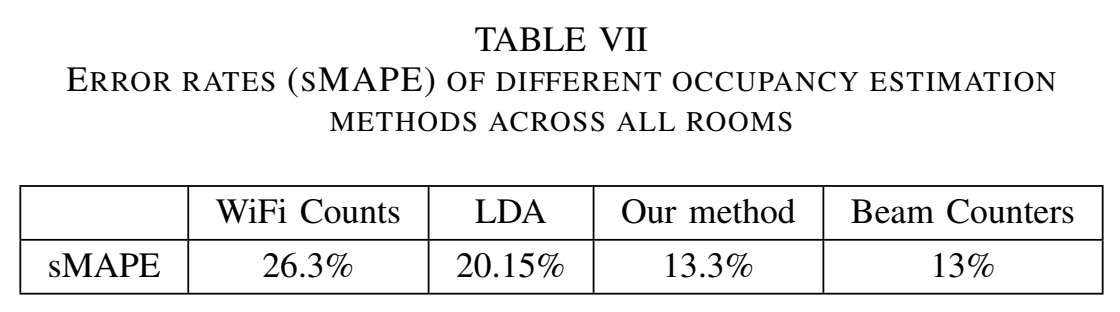


Berdasarkan tabel diatas, metode pada penelitian ini diuji pada kapasitas hunian yang lebih besar, dengan tingkat error lebih rendah dan tentunya tanpa biaya sepeser pun. Terlihat bahwa metode ini memiliki keunggulan dibandingkan dengan metode pada penelitian serupa sebelumnya.

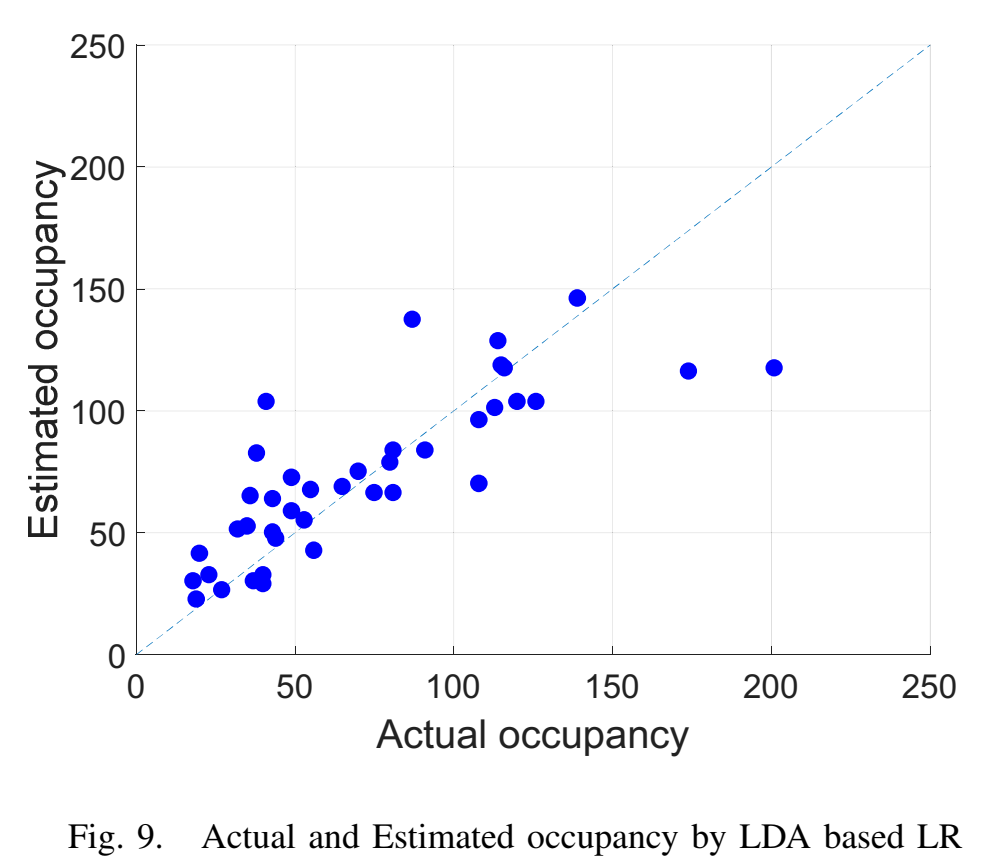
Bersamaan dengan penelitian ini pada ruangan yang sama sebagai objek penelitian dipasang sebuah EvolvePlus wireless beam counter untuk menghitung jumlah penghuni pada suatu ruangan [16]. Akurasi yang dihasilkan alat tersebut kemudian dibandingkan dengan klasifikasi LDA, klasifikasi LDA dengan LR, dan perhitungan mentah jumlah konektivitas WiFi. Tolak ukur yang dipakai adalah symmetric mean absolute percentage error (sMAPE). Formula pada perhitungan tolak ukur tersebut adalah :



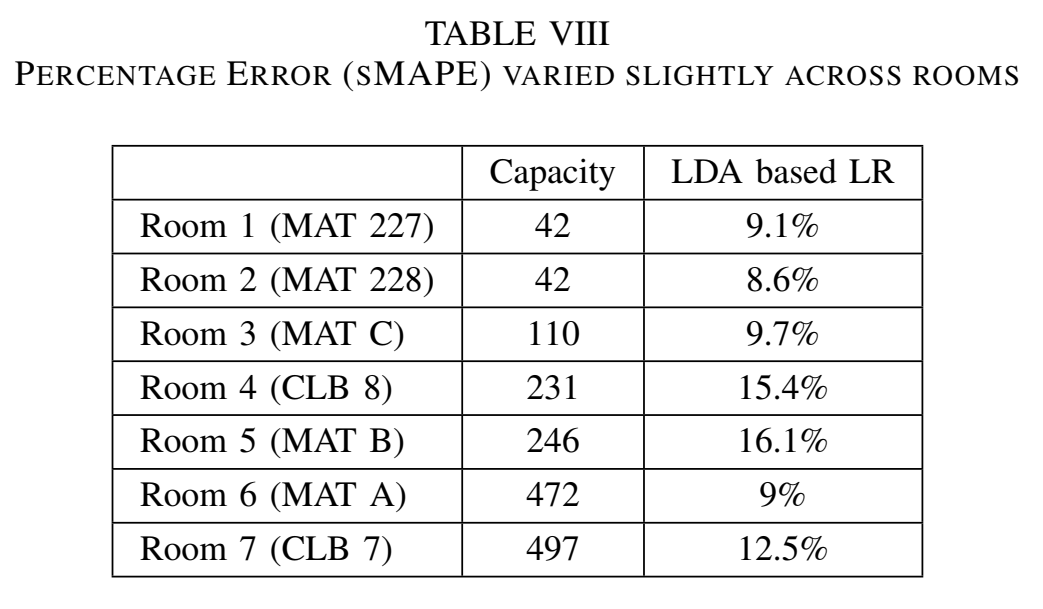
Dimana Ai adalah jumlah sebenarnya dan Fi adalah hasil peramalan dari input regressi yang berasal dari n input. Hasil sMAPE tertera pada tabel berikut :



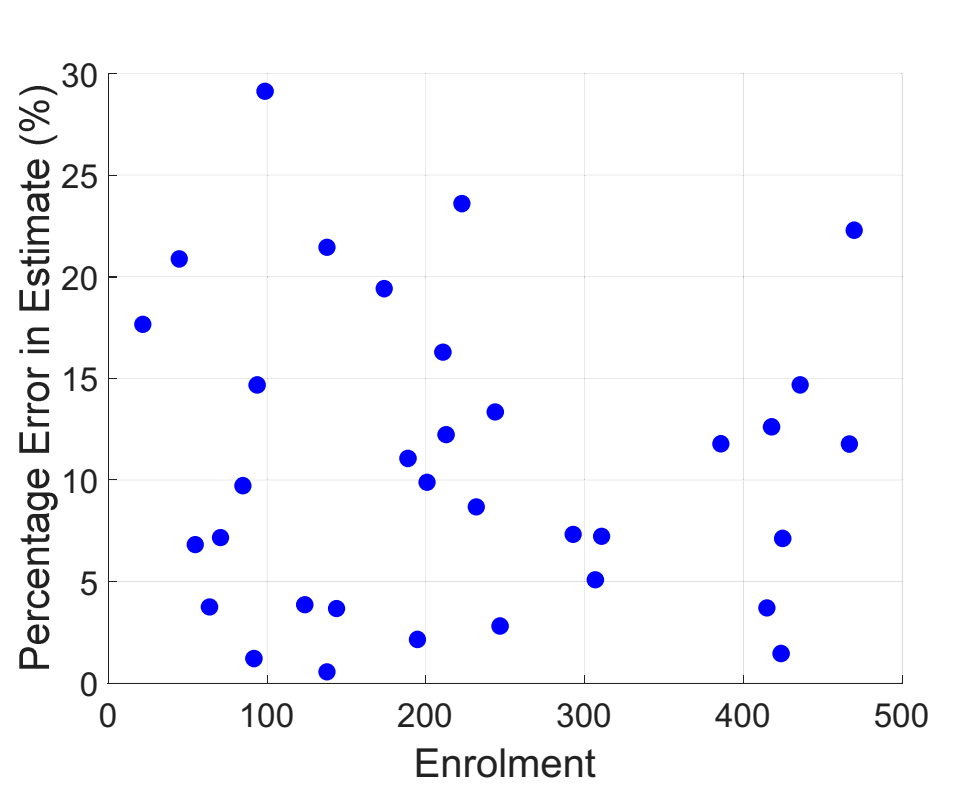
Pada tabel tersebut tertera bahwa metode yang digunakan pada penelitian ini telah mengurangi tingkat error hingga setengahnya terhadap perhitungan mentah koneksi WiFi. Dikarenakan data occupant berbaur dengan bystander dan penghuni yang tidak melakukan koneksi WiFi pada AP ruangan kelas. Sementara pada metode LDA data baru dipisahkan dari bystander saja. Tingkat error paling kecil ditunjukkan oleh hardware beam counter. Metode penelitian ini hanya berjarak sangat sedikit performansinya dengan hardware tersebut. Pada gambar dibawah ini dijelaskan bagaimana hubungan data dari prediksi yang dihasilkan oleh metode LDA dan LR dengan jumlah penghuni sebenarnya.



sMAPE kemudian digunakan untuk melihat performansi metode LDA dan LR dalam memprediksi tingkat hunian pada kelas yang ada. Range jumlah penghuni adalah dari 0-500, diterapkan pada kelas dengan estimasi kapasitas berbeda-beda yang dijelaskan pada tabel dibawah ini :

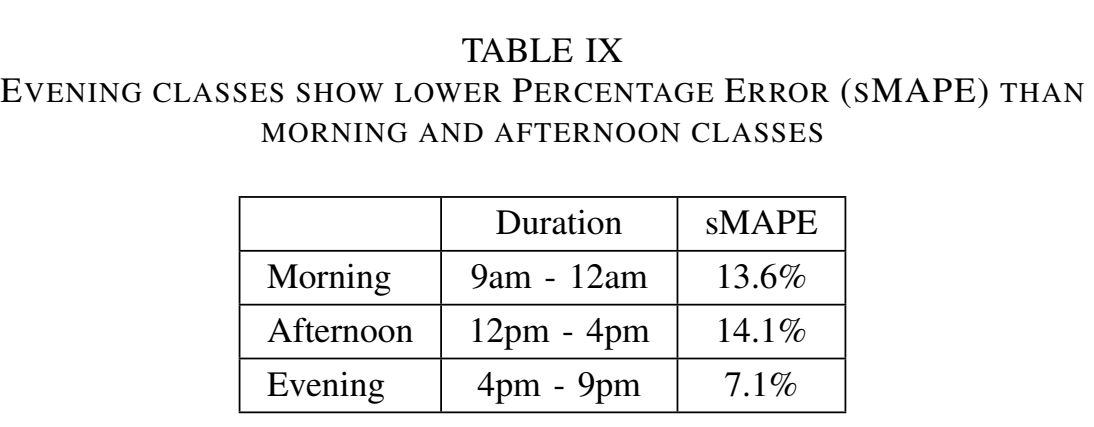


Distribusi sMAPE dengan jumlah mahasiswa terdaftar kemudian dibandingkan. Dan hasilnya terlihat seperti gambar dibawah ini :



Terlihat pada variasi nilai tingkat error pada setiap kelas beserta macam-macam kapasitasnya dan berbagai jumlah mahasiswa terdaftar. Metode ini dapat dilakukan pada seluruh kampus.

Posisi waktu kuliah juga perlu diperhatikan, terdapat kemungkinan bahwa kuliah di siang hari dan sore hari memiliki jumlah penggguna WiFi yang lebih sedikit. Nilai tingkat kesalahan prediksi pada waktu pagi, siang, dan sore hari harus diperhatikan. Berikut nilai sMAPE pada tiap-tiap waktu.



Terlihat nilai kesalahan meningkat pada siang hari, yaitu dari 13.6% di pagi hari menjadi 14.1% di siang hari. Akan tetapi pada sore hari tingkat kesalahan justru menjadi sangat kecil, yaitu 7%. Penurunan tingkat kesalahan ini dimungkinkan karena pengguna kelas pada sore hari adalah mahasiswa pascasarjana. Dimana mereka tidak berada di kampus dari waktu pagi hari dan hanya memiliki jadwal kelas di sore hari. Distribusi lokasi AP juga akan mempengaruhi kemampuan untuk memprediksi tingkat hunian. Beberapa lokasi memiliki kemungkinan tidak memilki AP sama sekali. Akan tetapi, pada lokasi dimana penelitian ini dilaksanakan, yaitu UNSW. Semua ruangan dilengkapi oleh konektivitas WiFi, bahkan dengan banyak AP terpasang pada satu ruangan.

Hasil dari penelitian ini dirangkum sebagai berikut :

1. Objek pengumpulan data untuk memprediksi tingkat hunian adalah data konektivitas WiFi yang tersebar pada beberapa lokasi ruangan kelas di kampus UNSW.
2. Data konektivitas WiFi perlu di saring untuk memperbesar akurasi estimasi tingkat huni. Hal ini dikarenakan data tersebut terkumpul tidak hanya dari pengguna yang berada di dalam ruangan kelas, tetapi juga berasal dari pengguna yang berada di luar ruangan kelas. Pengguna di luar ruangan kelas memungkinkan terkoneksi pada AP kelas karena jaraknya yang berdekatan. Sehingga pengguna dibedakan menjadi occupant (berada di dalam kelas) dan bystander (berada di luar kelas)
3. Klasifikasi data dengan metode LDA pada dua kelas, yaitu occupant dan bystander menghasilkan tingkat akurasi 84% untuk prediksi occupant dan 81% untuk prediksi bystander.
4. Tingkat kesalahan prediksi ketika hanya menerapkan metode LDA masih cukup besar, yaitu 20.15%. Untuk memperkecil tingkat kesalahan data occupant dimasukkan sebagai input pada metode LR, hal ini untuk menangkap jumlah penghuni yang tidak terkoneksi dengan WiFi. Gabungan LDA dan LR memperkecil tingkat kesalahan hingga 13.3% , hanya selisih sedikit dengan Beam Counter dengan tingkat kesalahan 13%.
5. Metode LDA dan LR dapat digunakan pada keseluruhan kampus dikarenakan nilai variasi yang kecil pada ruangan dengan kapasitas berbeda dan berbagai jumlah mahasiswa terdaftar.
6. Membuktikan bahwa WiFi soft sensor dapat digunakan untuk melakukan estimasi tingkat hunian pada ruangan dengan kapasitas besar sekalipun.

# Usulan Pengembangan Penelitian

Pemisahan antara bystander dengan occupant dapat dilakukan dengan tambahan analisa kebiasaan keduanya. Misalnya aktivitas selama koneksi, bystander mungkin tidak terlalu banyak melakukan browsing dikarenakan koneksi WiFi mereka terjadi tidak karena sengaja atau sinyalnya lebih lemah. Melihat tindakan yang dilakukan oleh bystander ketika ia tidak berada di ruangan tersebut [1], apakah ia akan melakukan koneksi dengan AP pada ruangan dimana ia tidak berada di dalamnya atau mencari AP yang terpasang di ruangan dimana ia berada. Bila ia cenderung melakukan koneksi dengan AP ruangan terdekat, apakah lokasi ia berada tidak memiliki AP . Dua fitur yang mungkin dapat diteliti lebih lanjut adalah :

* MAC address randomization
* RSSI threshold

Lokasi-lokasi terdekat baik ruangan, atau titik berkumpul luar ruangan yang tidak memiliki AP adalah potensi dimana banyak bystander berada. Tingkat kebutuhan mereka terhadap WiFi juga dapat di analisa. Apakah mereka sangat membutuhkan WiFi, sehingga mengambil koneksi atau tidak. Kemungkinan penerapan regressi dapat dilakukan terhadap keberadaan bystander juga. Sehingga jumlah data occupant dapat dipisahkan dengan lebih baik. Proses klasifikasi bystander dapat menggunakan klasifikasi untuk menentukan kemungkinan adanya ruangan berisi tanpa AP yang cenderung melakukan koneksi pada titik tertentu [2]. Dikarenakan tidak selalu ada list mahasiswa terdaftar pada suatu ruangan kelas untuk menyaring apakah ia benar-benar occupant atau bukan.

Mengenai occupant yang tidak melakukan koneksi WiFi perlu dilhat kecenderungan mereka untuk tidak melakukan koneksi. Bila mereka tidak memiliki perangkat maka faktor kepemilikan perangkat berperan, bila mereka memiliki koneksi internet mandiri, maka faktor kepemilikan koneksi internet berperan. Time series analisis juga perlu diperhatikan. Klasifikasi tidak hanya fokus pada bystander dan occupant saja, penentuan min dan max terhadap estimasi untuk menentukan apakah ruangan tersebut sedang terisi secara max atau berada di level minimal juga perlu diketahui.

Selain itu kemungkinan digunakannya sniffing untuk mendeteksi pancaran WiFi dan Bluetooth BLE, mungkin dapat dilihat sebagai perbandingan juga untuk menentukan akurasi klasifikasi [3] .

Referensi untuk pengembangan penelitian :

[1] G. Vanderhulst, A. Mashhadi, M. Dashti, and F. Kawsar, “Detecting human encounters from WiFi radio signals,” *ACM Int. Conf. Proceeding Ser.*, vol. 30-November-2015, pp. 97–108, 2015.

[2] D. Muchlinski, D. Siroky, J. He, and M. Kocher, “Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data,” *Polit. Anal.*, vol. 24, no. 1, pp. 87–103, 2016.

[3] E. Longo, A. E. C. Redondi, and M. Cesana, “Accurate occupancy estimation with WiFi and bluetooth/BLE packet capture,” *Comput. Networks*, vol. 163, 2019.

# Referensi Pada Paper yang direview

1] “Occupancy sensors market - industry trends, opportunities and forecasts to 2023, ”https://www.researchandmarkets.com/research/w4kck9/global occupancy?w=5, Dec. 2017, [Online; accessed 27-03-2018].  
[2] B. Dong, “Occupancy detection through an extensive environmental sensor network in an open-plan office building,” in Eleventh Int. IBPSAConference, no. January 2009, 2014.  
[3] Y. P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, and M. A. Little,“Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction,” Proc. of the 2016 ACM Int. Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ’16, pp.1016–1027, 2016.  
[4] D. Sgouropoulos, E. Spyrou, G. Siantikos, and T. Giannakopoulos,“Counting and Tracking People in a Smart Room : an IoT Approach,”in 2015 10th Int. Workshop on Semantic and Social Media Adaptationand Personalization (SMAP), 2016.  
[5] F. Paci, D. Brunelli, and L. Benini, “0, 1, 2, many - A classroom occupancy monitoring system for smart public buildings,” Conferenceon Design and Architectures for Signal and Image Processing, DASIP, vol. 2015-May, 2015.  
[6] Y. Yang, Z. Li, and K. Pahlavan, “Using iBeacon for Intelligent In-Room Presence Detection,” in 2016 IEEE Int. Multi-Disciplinary Conferenceon Cognitive Methods in Situation Awareness and Decision Support(CogSIMA), 2016.  
[7] G. Conte, A. A. Nacci, V. Rana, and P. Milano, “BlueSentinel : a first approach using iBeacon for an energy efficient occupancy detection system,” in 1st ACM Int. Conference on Embedded Systems For EnergyEfficient Buildings (BuildSys) 2014, no. Nov, 2015.  
[8] T. Yoshida, “Estimating the number of people using existing WiFi access point in indoor environment,” 6th European Conference of ComputerScience (ECCS ’15), pp. 46–53, 2015.  
[9] Y. Jiang, X. Pan, K. Li, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang,“ARIEL: Automatic Wi-Fi based Room Fingerprinting for Indoor Localization,” Proc. of the 2012 ACM Conference on Ubiquitous Computing, pp. 441–450, 2012.  
[10] N. T. Nguyen, R. Zheng, and Z. Han, “UMLI : An Unsupervised Mobile Locations Extraction Approach with Incomplete Data,” in 2013 IEEEWireless Communications and Networking Conference (WCNC), 2013.  
[11] R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen, “Measuring building occupancy using existing network infrastructure,” 2011 Int.Green Computing Conference and Workshops, IGCC 2011, 2011.  
[12] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal, “Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings,” Proc. of the 11th ACM Conference onEmbedded Networked Sensor Systems, p. 17, 2013.  
**[13] A. E. Redondi, M. Cesana, D. M. Weibel, and E. Fitzgerald, “Understanding the WiFi usage of university students,” 2016 Int. WirelessComm. and Mobile Computing Conference, IWCMC 2016, pp. 44–49,2016.**  
**[14] K. Akkaya, I. Guvenc, R. Aygun, N. Pala, and A. Kadri, “IoT-based occupancy monitoring techniques for energy-efficient smart buildings,” 2015 IEEE Wireless Communications and Networking Conference Workshops, WCNCW 2015, no. March, pp. 58–63, 2015.  
[15] Y. Wang and L. Shao, “Understanding occupancy and user behavior through wi-fi-based indoor positioning,” Building Research & Information, vol. 0, no. 0, pp. 1–13, 2017.  
[16] T. Sutjarittham, H. Habibi Gharakheili, S. Kanhere, and V. Sivaraman,“Data-Driven Monitoring and Optimization of Classroom Usage in a Smart Campus,” Proc. of the 17th ACM/IEEE Conference on InformationProcessing in Sensor Networks (IPSN 2018), 2018.**