Abstract. Meeting rooms are often sparse for companies and managing their usage is a difficult process. Also, in big companies it is important to find which teams interact with each other in order to optimize spatial localization. Based on data provided by real companies, this project tries to address these issues using machine learning algorithms trained with data received from sensors in meeting rooms and data obtained from Wi-Fi access points placed in companies buildings. Several previous studies in these fields are reviewed. Solutions using machine learning algorithms, data protection measures and visualization techniques are proposed, followed by evaluation analysis.

Keywords: Machine Learning · Internet of Things · Visualization · Data Science · Management · Occupancy · Forecasting · People Flow.
Table of Contents

1 Introduction ................................................. 3
  1.1 Available data ........................................ 3
  1.2 Work goals ............................................. 4
2 Related work .................................................. 4
  2.1 Room occupancy ......................................... 4
  2.2 People flow ............................................. 6
3 Solution proposal ............................................ 7
  3.1 Calibration process ................................. 9
      Detecting room occupancy .............................. 9
      Predicting room occupancy ............................ 10
  3.2 Prediction process .................................... 11
  3.3 Flow process ......................................... 11
  3.4 Visualizations ........................................... 12
      Room occupancy prediction ................................... 13
      People flow ............................................. 14
  3.5 Data identification and anonymization .......... 15
4 Evaluation .................................................... 17
  4.1 Prediction models ..................................... 17
  4.2 People flow ............................................. 17
  4.3 Visualizations ........................................... 18
5 Work scheduling .............................................. 18
6 Provided data and log extraction ................... 19
7 Conclusion .................................................... 20
A Appendix A ................................................... 23
1 Introduction

For every company there is a constant need for efficient use of physical space. This is specially true for meeting rooms inside the buildings which can be booked at any time by different employees, making the task of managing resources a difficult one. In big companies there are often times where the demand for meeting rooms exceeds the capacity of a workplace building, leaving workers without a proper meeting place. Big companies also have multiple teams which might often interact with each other. Having such frequently interacting teams located in a closer geographical location will reduce the time workers spend moving inside the company and may also result in better performance overall [24]. This project aims to address the two points mentioned above using technology methodologies in the fields of internet of things, machine learning and information visualization.

1.1 Available data

A big Portuguese company has provided data regarding its workers’ meeting bookings and people flow inside the company’s facilities. Data regarding meetings was collected using the internal meeting booking system (Microsoft’s Office 365), sensors placed in the meeting rooms and tablets located at the room’s entrance/building lobby.

The basic information given by the booking system includes:

- Meeting subject;
- Organizer;
- Number of participants;
- Planned start/end;
- Real start/end;
- Meeting room;
- If the meeting was quick booked;
- If the meeting ended early.

The company’s employees are able to book meetings using the Microsoft Office Outlook software using their computers and must check-in/check-out through the tablets placed at the entrances whenever they use a room for meetings. The tablets also allow for a room to be quick-booked by an employee directly if there are no meetings scheduled at that particular time.

The sensors placed in each room allow the extraction of more detailed information about how the meeting rooms are being used, such as:
The equipment has been tested thoroughly by the company’s engineering team and is considered to be accurate and therefore reliable for modelling purposes. People flow inside the company will be analysed using data obtained through the various Wi-Fi access points placed inside the facilities and includes the devices which were connected to an access point at a given time.

It should be noted that the provided data’s quality will be an important factor for the success of any predictive model / visualization obtained through this project.

1.2 Work goals

The purpose of this project will be to:

- Develop a predictive model to determine whether a room will be occupied or not in a future time;
- Determine how many rooms will be necessary in a specific time;
- Estimate the likelihood of an employee booking a meeting room and not using it;
- Find how rooms are being used outside of meeting bookings;
- Study the flow of people through Wi-Fi access points connections in order to optimize teams positions;
- Develop visualizations so that the data can be easily interpreted by any individual;
- Integrate the solutions with existing database systems;
- Ensure compliance with the European General Data Protection Regulation (GDPR) [12]
- Evaluate accuracy of the solution.

Integration of the predictive models in existing booking systems inside companies and their impact in room availability represents a complex process itself and therefore it will not be addressed in this project.

2 Related work

2.1 Room occupancy

There have been several studies related to room occupancy predictions, namely for energy saving optimizations [11] [23]. However most of this
work is focused on space occupancy as a whole without taking into account previous bookings made by individuals.

This section will show several known machine learning algorithms which might be useful in predicting whether a room will be occupied or not at a given time. These algorithms were discussed in occupancy detection problems and can be adapted to respond whether a room will be occupied in the future or not by using different features.

One study on room occupancy classification refers to a number of different algorithms which can be used to classify whether a room is occupied based on different sensors inside rooms [7]. The dataset used by this article was built using data from sensors (temperature, humidity, light, CO₂ levels and humidity ratio) along with features related to the time that passed since the midnight of a specific day and a boolean value referring to whether the current day is or isn’t a weekend. The authors tested different algorithms with distinct features selected in order to evaluate the accuracy of each model and features pairs. The algorithms evaluated were the following:

- Random Forest [3];
- Gradient Boosting Machines [13];
- Linear Discriminant Analysis [8];
- Classification and Regression Trees [4].

In general all the algorithms showed high accuracy scores for their predictions, being able to determine whether a room is occupied or not with success.

Another study tries to address this problem in the hospitality industry [6] in order to predict the demand for hotel rooms. Several regression machine learning algorithms are detailed which can be used to predict room occupancy:

- Ridge Regression [15];
- Kernel Ridge Regression [18];
- Multilayer Perceptron [21];
- Radial Basis Function Networks [5].

This article details important steps in data pre-processing for training models containing time series. In order to predict if a room’s occupancy status in the future, lag variables [9] are created for each instance of the dataset. These lag variables correspond to n previous room observations in specific time frames (for example, a day ago, a week ago, a month ago) which are introduced in the dataset along with current observations (see
Table 1 for an example). This allows the model to predict a instance in the future while those $n$ previous room observations are available. From there forward one can repeat training the model with already predicted data so that it is possible to look further into the future.

**Table 1.** Example of a dataset containing lag variables. The second column *occupants* represents the number of people inside a room in the period given in the first column. Columns *occupants1*, *occupants2* and *occupants4* represent the values of *occupants* observed in past periods, in this case one, two and four hours behind, respectively. The arrows presented represent the lag variables source for the last entry of the dataset. This data is meant for exemplification purposes and does not represent real world observations.

<table>
<thead>
<tr>
<th>time</th>
<th>occupants</th>
<th>occupants1</th>
<th>occupants2</th>
<th>occupants4</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10:00</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11:00</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>12:00</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>13:00</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Another important pre-processing step mentioned is the inclusion of features specific to the time when previous instances were observed, such as month of the year, day of week, holidays or if the instances were observed during tourism season. This can be relevant for the meeting bookings since they are dependent on the time they occur (for example, almost everyone should be out of office during August due to summer holidays, Christmas and other events).

Additionally the authors took into account room reservations when building one of their datasets, which is important for accurate usage prediction for meeting room bookings. Their results show that the inclusion of extra features mentioned above improved their models are resulted in less error overall.

### 2.2 People flow

There have been various studies in people flow analysis, although not specifically for companies indoor tracking. This section will go briefly through existing work and examine the areas that might be useful for studying people flow inside the company.

Crowd densities and pedestrian flows in an airport have been studied in [22]. This article shows that it is possible to obtain information about people flow using Wi-Fi (and Bluetooth) systems, which can be
used to reduce movement times within a determined space. The different approaches followed to address this matter are as follows:

- **Naive approach** where the algorithm only counts the number of unique MAC addresses in two different spaces during a specific time-frame. This solution can cause a number of false-positives for people movement and the direction cannot be determined;

- **Time-based approach** in which the time when a MAC address was registered for the first and last time at each access point is considered for determining the direction of movement;

- **RSSI-based approach** where the received signal strength indication (RSSI) is taken into account in order to reduce the false-positives caused by overlapping of access points.

- **Hybrid approach** which considers both the RSSI value and the time of when a MAC address was connected to an access point.

These approaches were concluded to be accurate by the authors and can all be applied to detect patterns in companies employees movement inside buildings in this project scenario using Wi-Fi access points.

The article in [1] details the analysis of movement for a big quantity of data relating spacial positioning of a car. This is not directly related to this project since the type of data is different, however from the analysis point of view it suggests that visualization techniques might be the best way to analyse movement data. The authors start by using a representation of the physical space where the movement happens (a map) and draw patterns over it. They also use tables with locations and symbols to simplify reading and interpreting the data.

Another article studying people flow focuses on the movement of individuals inside a university campus [14]. The authors chose to analyse Wi-Fi access point signals through a visualization platform, again suggesting that viewing data regarding people flow is of significant relevance for a good analysis of the data.

### 3 Solution proposal

The solution proposed for this project will be divided in four different parts. A diagram showcasing how each part is connected along with a brief description for each one can be seen in Fig. 1.
Fig. 1. Diagram describing the solution's flow of data from sensors/access points until the final visualizations and steps for processing data.
3.1 Calibration process

The calibration process will be the first step in room occupancy prediction. In this step the data received from the company’s databases will be assembled and processed into a single dataset.

The idea is to create two different predictive models: one for detecting whether a room is occupied or not and another for predicting room occupancy in the future.

The algorithms which will be tested are referred in section 2. For implementation two programming languages were considered, R and Python, since both provide many packages with machine learning models and ways to visualize data. Python was chosen due to familiarity with the language and more flexibility for treating data. The implementation will be done using the functions included in scikit-learn package [20] for machine learning algorithms, pandas [17] for dataset building and others that might be useful to generate derived variables such as datetime and holidays. Information about the algorithms accuracy should also be generated with simple charts created using Python’s matplotlib package [16] for easier accuracy analysis.

All datasets used in the models will be split in three sub-datasets for training, validation and testing to ensure accuracy and avoid over-fitting.

The calibration should only be done in long intervals of time, for example weekly or monthly, since it is a computationally costly process. The models will be saved to disk so they can be used for the prediction process later on.

It should be noted that the data available from the company is limited in terms of time since the data is only being collected for less than a year. Still, the solution proposed will take into account that there will be enough data to feed the models as to reflect seasonality seen through many years of data.

Detecting room occupancy The model for detecting room occupancy will be built using the sensor data referred in section 1.1 as features and meeting check-ins or their absence as class for each room reservation. This will be done so it is possible to classify a room as being occupied or not even when in the absence of reservations, in order to address cases when rooms are being used for unregistered meetings or for private phone calls, for example. The dataset used to train the model will not contain any time related features.

The training using feature selection should also give insights of whether sensor data such as the infrared thermal sensors for counting people are
descriptive enough so that they can be used as a single attribute capable of determining to whether a room is occupied or not without a predictive model.

**Predicting room occupancy** Using the model discussed above it is now possible to create a dataset that is able to predict a rooms’ occupancy at a given time. The dataset must correlate meeting bookings, the occupancy status of the room and extra information regarding the date/time of the instances.

Since this data refers to temporal events a good way to address the periodicity will be to create an instance for each half hour of a day and then fill the following features:

- Day of the month;
- Month;
- Year;
- Weekday;
- Is it a weekend?;
- Is it a holiday?;
- Are there holidays in a 7-day period ahead/behind?;
- Season;
- Day of the year;
- Weekday;
- Half hour (1 to 48);
- Room;
- Booking status;
- Organizer (if booked);
- Occupancy status (class);
- Occupancy status lag variables.

These features should be enough to create an accurate model to predict room occupancy. Feature selection will be used to determine the combinations of features which result in models with higher accuracy. It should also be analysed whether including booking status is beneficial or not for the problem since it might bias the results, making the algorithm always predict that a room will not be occupied if no booking is planned for it. The Organizer feature might also need to be factorized so it works well with all the proposed algorithms in section 2.
3.2 Prediction process

The prediction process is where the calibrated model will be used to predict each rooms occupancy status at future periods of time.

Starting with the time period for which there isn’t information about the rooms’ occupancy status (a time in the future), the algorithm will start building a dataset containing the features used to calibrate the model as described in section 3.1. The number of entries in this dataset will be limited to whether or not its lag variables are available (see table 1).

The algorithm will then use the generated dataset to classify the rooms occupancy status. These results might be used to generate yet another dataset since they allow the input of more lag variables for adding entries to predict, making it possible to look further into the future.

The prediction process should be executed daily, updating the predicted values with the observed ones and starting building new datasets as described above.

After the prediction process is finished, predicted instances should be saved to disk in CSV (comma separated values) format so that they can be visualized as described in section 3.4.

3.3 Flow process

The flow of people will be analysed based on data received from Wi-Fi access points present in the facilities, using the methodologies described in section 2.2.

This part of the project is very dependent on data quality for obtaining good results. Since Wi-Fi connections are considered to be very inconsistent (devices might disconnect arbitrarily due to loss of connection and users can turn off their device’s Wi-Fi at will) results might not be accurate and adaptations could be necessary to clean the data and make sure it produces accurate results. It is also important to have knowledge about the building’s plant to correctly analyse the produced information.

The data regarding movement should be presented through a visualization, described in section 3.4. Data generated from people flow should contain:

- **Average daily/weekly working hours** using the time individuals spent connected to any access point;
- **Popular times when people disconnect from the network** which can be used to analyse how individuals are taking breaks (for example, at lunch time);
– **Popular trajectories for getting from point A to B** which can point to teams which work together often;
– **Find the periods in which a higher density of people is found inside the spaces**;
– **Find the types of devices employees are using to access the web**.

None of the information gathered should point to a specific individual as to comply with GDPR ruling [12]. See section 3.5 for more details.

### 3.4 Visualizations

Visualizations must be implemented for this project to ensure that individuals with low technical skills will be able to analyse the data produced in previous sections. With this in mind, a good way to develop a visualization which can be accessed by any individual is to use a platform most people already use in a daily basis: the browser. For this reason, the proposed solutions will be web-based, using technologies which can be loaded in a browser.

For this project several visualizations will be developed from the ground up using Javascript, a scripting language which is widely used for web programming. There are many Javascript frameworks available which can be used for this project’s purpose.

A popular framework for building data related visualizations is the Data Driven Documents framework *d3.js* [2]. It packages a whole set of tools which can be useful not only for making charts and graphs, but it can also be used to represent moving particles over graphics\(^1\).

Several tools which could be useful for representing indoor building plants were analysed. *maptalks*\(^2\) provides tools for working with maps, however upon trial it proved to be too difficult to use when representing anything which wasn’t a real-world map. *three.js*\(^3\) is a tool used to develop video games which could be used to create a minimal representation of the building’s indoor physical space, however it is difficult to code for and might introduce increased complexity for visualization due to its camera components.

Because the solution requires so many specific details, the final decision is to use *d3.js* to build from scratch for all the points of this project’s visualizations.

---

\(^1\) See [http://www.decembercafe.org/demo/plane/](http://www.decembercafe.org/demo/plane/) for a visualization of airplane connections using *d3.js* (last opened in December 6, 2019)
\(^2\) [http://maptalks.org](http://maptalks.org) (last opened in December 6, 2019)
\(^3\) [https://threejs.org/](https://threejs.org/) (last opened in December 6, 2019)
Below different visualizations for each part of the project will be discussed.

**Room occupancy prediction** For room occupancy information a single web page will be used as a visualization tool. There will be a calendar in which users can select a day with available predictions (see Fig. 2) and a schedule for selecting a time frame (by half hours) (see Fig. 3). Below the calendar and schedule there will be a list of rooms accompanied by relevant data (see Fig. 4). The rooms will be presented with different colours (green for free, red for occupied). The rooms are clickable and information about the selected room’s occupancy will be shown in the page. When hovering the mouse over a room it, a tooltip describing the room’s status for the day will appear. Information about occupancy percentages should only take into account working hours, which will be defined after analysing the dataset on user flows to determine the times where most people are present in the company.

![Fig. 2. Depiction of the calendar used to select days for accessing prediction information on room occupancy. Arrows on top can be used to navigate through months. Only a day can be selected at a time.](image)

Additionally aggregators for multiple days could be added for even more detailed information (for example, studying room occupancy during the summer).
This visualization should contain real and predicted data depending on the date/time selected in the calendar, with this information being displayed to the user.

**People flow** The building being analysed will be drawn using the Javascript framework *d3.js*, using several layers to represent different floors. The building plant will be modelled using JSON (JavaScript Object Notation) using simple points coordinates and floors to point how the plant should be drawn.

The page will have a similar calendar and schedule as seen in the room occupancy visualization mentioned in the previous section. The page will then show a floor with its respective spaces and information about the number of connected devices (see Fig. 5).

Movement should be represented using dots that move along the different rooms in the building, representing people. Buttons similar to those of a video player will be placed under the building visualization allowing to start playing movement, speeding up or slowing down the animation. Current time should be updated in a information panel located next to the plant.

The dots should have a colour associated with the room they were placed in before hitting the play button. New dots appearing during animation take the colour associated with that room. It will be also possible to track a dot’s location history by showing a tooltip when hovering the mouse over it.
Default animation speed should be slow enough so that patterns can be easily identified visually. In order to visualize more than one floor there must be an option available to select multiple floors to be shown in the visualization simultaneously, side by side. Additionally, a option to follow a dot by clicking on it will be included, so that the floor changes automatically when an individual moves from one floor to another.

Information about average daily/weekly working hours, popular times when people disconnect from the network, popular trajectories between two different points and periods with higher amounts of people inside the facilities should be shown next to the building visualization.

A bar chart will also be included to show the different types of devices employees are using to access the internet inside the company.

### 3.5 Data identification and anonymization

Current European regulations imposed by the General Data Protection Regulation (GDPR) [12] prohibit the usage of personal data which can be used to identify individuals without consent. The identification is not limited to direct links between data and individuals, such as names or MAC addresses, but also aggregates of information which may point indirectly to an individual [19]. Measures must be taken to anonymise the personal individuals’ data in order to avoid legal repercussions.
When using attributes relative to the room bookings’ owners to build the datasets for machine learning, any names or possible links with existing database systems will be anonymised using (symmetric) encryption with a private key which should not be shared/used by anyone other than the system. Machine learning models will be able to produce the same results as they would with the original, unencrypted data, making this a simple solution which will prevent user identification. Any other type of indirect linkage to individuals through system-generated datasets should not be possible, since:

- The datasets should be destroyed after training the model and obtaining results;
- The prediction dataset only contains each room’s occupancy status for time periods, not containing specific individual data.

Room occupancy predictions information can still be crossed with the meeting bookings made within the company. For example, one can infer that an individual with bookings which are always predicted by the system as not happening isn’t occupying any of the rooms he/she books. However, this type of information is too specific and might not be observed in real world data.

For people flow analysis, there are two important things when considering individuals privacy. First, MAC addresses observed by the access points may link directly to an individual, which can be solved by anonymising each address when extracting them from the company’s system through encryption. Second, the physical location of a device within the facilities can be easily linked to the individuals through their workplace. For example, if subjects A, B and C work at room R and in the peoples flow visualization there are three points in room R present throughout most of the working hours, it would be easy to see that those three points belong to subjects A, B and C. In order to avoid people tracking, the points location within a room should be randomized each time the visualization is loaded, ensuring it is room-oriented and not individual-oriented.

Despite the details mentioned above, the company should still consider giving consent forms to individuals working/staying in the facilities detailing what is being collected and how is the data being processed.

---

4 Advanced Encryption Standard (AES) [10] is proposed as it is the algorithm adopted by the USA government and has few impact performance-wise.
4 Evaluation

This project will contain different pieces which must be evaluated separately. The sections below detail how testing is planned for each one of these steps.

4.1 Prediction models

The only way to evaluate a predictive model is to compare the predictions against what was verified in reality. This will be difficult for the use-case presented as that would require having a human being verifying if a room was in fact occupied for each half hour. Because of this, full accuracy data cannot be obtained.

However, for booked meetings, accuracy can be determined by comparing predictions against check-ins using the tablets placed in the facilities. This will allow a partial evaluation of the system.

In order to evaluate the predictions, predictive data corresponding to a full month will be gathered and after the month has passed, data regarding booking check-ins will be extracted from the company’s system. Then both data will be compared in order to produce an accuracy value (the number of correct classifications divided by the total number of classifications). Other measures will be used to evaluate the model such as root mean square error (RMSE) and mean absolute percentage error (MAPE).

4.2 People flow

Evaluating people flows in a trajectory point of view will not be possible since there are currently no other tracking mechanism implemented that can be compared with the Wi-Fi access points method. For most buildings, the only information available about workers is the time of arrival and departure from the facilities, which is not enough data for evaluation. Therefore, this can be left as a future work for the company if they ever decide to implement a tracking mechanism inside the company, such as using NFC for opening every single door inside the building.

Some of the data can be evaluated, such as the average working hours and popular times when people disconnect from the network by comparing the data against employee arrivals and departures from the building, however such data might not be available for this project.
4.3 Visualizations

Visualizations can be evaluated through usability testing with different individuals, taking special attention to the ones with fewer technical skills.

The testing should be performed by giving users tasks which require exploring the visualizations and asking a predefined set of questions referring to:

- Ease of use;
- Utility;
- Design;
- Satisfaction.

The tasks should be evaluated based on the time of execution and user errors. Any changes to the visualizations should be considered in order to reduce time of execution and user errors.

5 Work scheduling

The work scheduling proposed for elaborating this project can be viewed in Fig. A.1. The scheduling starts with the second semester and was planned taken into account previous experience with projects done in similar subjects.

The first three weeks will be dedicated to analysing available data and building the datasets which will be used for creating the predictive models and then try the different algorithms proposed and evaluate their accuracy, trying multiple feature combinations in order to reach the best solution while avoiding overfitting. Relevant charts for the results will also be developed.

For the next three weeks the peoples flow inside the company will be analysed in order to model the data so that it is possible to find patterns and get the information discussed in section 3.3. This task will probably prove to be difficult due to the inconsistencies which Wi-Fi tracking might bring to the datasets.

Building visualizations from scratch will be the most time-consuming task since it will require writing code from scratch and it will take up to a month, if not more, to develop.

Finally, evaluation will take place in the first two weeks of May and the semester remaining time will be spent writing the dissertation.
6 Provided data and log extraction

Throughout the semester some data was made available to ensure all the requirements for this project were met.

Data regarding room bookings and sensors was shared through Microsoft’s Power BI Desktop (PBIX) files, containing more than 70 tables which could be copied to Microsoft’s Office Excel and then saved as CSV files. Connections between different tables will be done using Python’s pandas [17] package in order to unify all data in a single dataset. Details about each table were recently shared and are being analysed.

For access point connections there wasn’t any database table available from which data could be extracted directly. Instead, logs from a company’s network server were extracted, which contained millions of entries representing events related to device connections in the network. A sample of the data extracted can be seen in Fig. 6.

Upon analysis three different events related to users’ connections were observed: connected, roam and disconnected. For every event a user MAC address is present along with other types of information.

- **connected** contains the access point’s address to which the user connected to. These events refer to when a user was seen in the network when not in a connected state (when disconnected from a previous session or seen for the first time in the network);
- **roam** contains details about the source and destination access point’s addresses. These events refer to when a user roams inside the company. Roams can be identified for either an access point or a channel change within one;
- **disconnected** contains the access point’s address from which the user disconnected from. The connection time and size of the data flow while connected is also present. These events refer to when the user was disconnected from the network, which means the MAC address is no longer connected to any access point in the system.
In addition to user-related events, others could be used to infer access point’s names which were describing enough to understand where in the building they were located.

A Python algorithm was built in order to get the important information from the log files, searching for lines with keywords and parsing them, obtaining the relevant data and saving it in CSV format. Lines containing \textit{roam} with channel changes only were dropped since they are not relevant for this project. The final data representation can be seen in Fig. 7.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
\textbf{date} & \textbf{time} & \textbf{user} & \textbf{status connected} & \textbf{from_ap} & \textbf{to_ap} & \textbf{bytes} & \textbf{from_ap_name} \textbf{to_ap_name} \\
\hline
20/11/2018 10:18:16 & 0:04 & roam & 44 & 46 & 0:04 & 2e & AP-RC & AP-RC \hline
20/11/2018 10:33:45 & 0:04 & roam & 44 & 46 & 0:04 & 2e & AP-S Dto & AP-Galeria \hline
20/11/2018 10:51:10 & 0:04 & roam & 44 & 46 & 0:04 & 2e & AP-Galeria & AP-S Dto \hline
20/11/2018 10:59:34 & 0:04 & roam & 44 & 46 & 0:04 & 2e & AP-S Dto & AP-Galeria \hline
20/11/2018 14:17:18 & 0:04 & roam & 44 & 46 & 0:04 & 2e & AP-Galeria & AP-S Dto \hline
20/11/2018 14:29:06 & 0:04 & roam & 44 & 46 & 0:04 & 2e & AP-S Dto & AP-Galeria \hline
20/11/2018 14:38:30 & 0:04 & disconnected & 44 & 46 & 0:16m & 26.53M & AP-RC & N/A \hline
\end{tabular}
\caption{Screenshot of the information extracted from the log files. Sensitive data has been blurred out.}
\end{table}

The information retrieved from the log files appears to be complete and detailed enough so that it can be analysed in the future. As seen in Fig. 7, it is possible to find the path an individual takes throughout the day. Some interesting information can be inferred from this data such as working times, lunch time and whether a device is mobile or not.

7 Conclusion

This report analysed several ways in which machine learning, IoT and visualization can help with the task of analysing data about meetings and people flow inside buildings.

Related work shows promising results when working with this type of data. However, machine learning algorithms represent distinct results for different datasets and there will be several challenges ahead before finding the best model for the described use-case. Also, analysing people flow with such variable data will be a challenging task and the visualization ideas proposed might need changes to meet the defined goals.
References

A Appendix A

Fig. A.1. Future work schedule.