**First meeting (Group) – 29-12-17, EC1-01**Detailed Structure of the dissertation and key areas - Topic discussed – Research areas and topics to research on listed – Project proposal explained – Sample project proposals emailed – for a cleared idea of what was expected of us.

**Second Meeting (Extended topic of research) – 11-01-18, Techno Centre**further explanation and breakdown, narrowing down the problem

**Third meeting (Project proposal -Literature Review) – 06 -02-18, William Morris:**Questions related to the project proposal content. How to address that were still uncertain like dataset and algorithms.

**Fourth meeting (Final Project Proposal Draft) 07-02-18, ASG-31:**Project proposal drafts and queries sent and feedback given via emails.

**Fifth meeting (Ethics, Data set) – 20-02-18, ECM-15:**Ethics application submission discussed and submitted. The initial public Kaggle based dataset of 30 households discussed and played around with. The option to use Davis’s electricity dataset arose. The dataset was received via email. The ethics application was updated as a public open source data was not being used. A consent form between the owner and researcher.

**Sixth meeting (Dataset Methodology) – 07-03-18, ASG-31:**The collection of the dataset explained. The manual techniques and the equipment required for the collection of the electricity consumption data as well as the pings that recorded the availability of the person at home. Ethics consent form designed discussed and the relevant elements and matter pondered over.

**Seventh meeting (Clustering, k-Means Algorithm) – 22-03-18, Planet Earth:**The split in the dataset discussed. The need for the split to be randomly picked not containing consecutive instances in the training data (70%) and the remaining consecutive occurrences to formulate the testing set (30%). The Kmeans algorithm and its nature discussed. The fact that the number of clusters (K) needs to be user provided.

**Eighth meeting (Deviation, Qualitative and Quantitative, Stats, Algorithms) – 27-03-18, EC Building:**Explanation that the understanding of the data machine learning had led me to understand that clustering alone is a stand-alone method to group data based on similarity and trends and an additional methodology needs to be implemented in order toclassify an input testing data for prediction. A method such as regression etc. The possibility of applying more than one algorithm and performing a statistical analysis to prove which works best in the given scenario. The possible ways to analyse the results, precision, accuracy, recall, f1 score etc discussed.

**Ninth meeting (End of term Presentation) – 29-03-18, William Morris:  
Attendees:** David Croft (Supervisor), Matthew England (Second Marker), Self.

All the progress in numerical values. 20% of the writeup along with nearly 45% of the coding done. What was left to be done, what were the problems encountered. Dissertation writeup breakdown, Word limit breakdown, project explanation from the bottom up. Its application and methodologies discussed.

Feedback from the presentation recorded on phone using voice recorder with consent of all present parties. Feedback included not to stress on the word limit too much, the necessity of having a rigorous statistical analysis in place, not just analysing the algorithms using built in functions like mean square error and precision score. The use of False negatives, false positives, true positives and true negatives to calculate precision and recall to ultimately crown 2 regression and classifier algorithms out of the implemented 4. The need of a personal reflection on the writing experience what went right? What went wrong? Etc. Write-up stance to be formal and in third person. Avoiding citations from websites and blogs. To clearly highlight the intended application and the actual implementation of the project as it the project title and explanation was prone to confusion as to what exactly was being implemented.

**Eleventh meeting (Draft check, run through) – 17-04-18, EC Building:**Limited feedback given at this stage as per university regulations. Suggestions and pointers as to what section might need another look etc. Structure, reflection and analysis section queries cleared out. Code unrequired but maintained in GitHub. Something of immense importance to be added to report itself otherwise snippets of relevant information to be added to appendix and referred to in the section of choice.

**Twelfth meeting (Draft check, run through) – 25-04-18, EC Building:**Analysis report, table of F-measure and statistical calculations. Feedback to keep it to the point and explain the findings and statistical calculations. Weekly log report, project management, analysis and project proposal, appendix, declaration of originality structure and placement discussed.

**Signed:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
David Croft   
(Supervisor)

**Dated:** \_\_\_\_\_\_\_\_\_\_\_\_\_28-04-18\_\_\_\_\_\_\_\_\_\_\_

**Project Management**

This project implemented the iterative nature of agile techniques that were inculcated in me during the entire course of this degree. Project, and time management has played a huge part in the success of this dissertation. The initial breakdown of tasks wasn’t ideal and exhaustive. It lacked many areas that showed themselves later at execution stage. The iterative adaptation and delivery of the project made it possible to go back to these areas and work on the accordingly.

Even in coding, after research based on the algorithms was carried out, all different algorithms (Neural networks, Kmeans clustering and Linear regression) were implemented and written from the bottom up, meaning their most basic initial stage and worked up in the form of reasonable increments. This helped breaking down a seemingly large problem from to smaller SMART (Specific, Measurable, Achievable, Relevant, Time limited) goals and tasks.

|  |  |
| --- | --- |
| **Sequential Technique** | **Iterative process (adapted technique)** |
| 1. Implement the data mining algorithm to predict occupancy of a house | 1. Import resources    1. Import Dataset    2. Import relevant libraries    3. Extract variables from dataset    4. Write the basic algorithm |

**\*** For instance, the above table shows the implementation of the algorithms using a sequential/un-planned technique as opposed to the adapted iterative technique with measurable distribution of work load.

The initial Gantt Chart included a list of tasks and the time they would take (in weeks) was heavily underestimated. The data set search took longer than anticipated

**Gantt chart Initial and Final**

**Sprint backlog** – with a footnote to explain all what’s happening (from February 9th onwards)

With graph.