300/303COM Detailed Project Proposal

First Name:	SHAHZEB
Last Name:	DAWOOD
Student Number:	7015511
Supervisor:	DAVID CROFT

SECTION ONE: DEFINING YOUR RESEARCH PROJECT

1.1 Detailed research question

Help: Your detailed research question is the statement of a problem within the computing domain which you will address in your project. Refining the research question involves narrowing down an initial question until it is answerable using a primary research method(s) that you will conduct during the time of your project. The refined research question must not be so general that it is answerable with a yes or no answer. It must not be so broad that you would be unable to achieve a solution during your project. The key to this is BEING SPECIFIC: Narrow down the method or technology you will use, narrow down the group that the question refers to (localize a general question) If the project is still 'too big', can you think of a way to work on a part of the problem? Avoid using words that cannot be measured, by you, without a huge research budget e.g. 'effects on society', 'effects on business'. *Example:* The initial question "Does cloud computing effect business" needs narrowing down *(for a start the answer is yes) W*hat is meant by cloud computing? Or 'effect'? Or 'business', in this question? Refining this first question will involve narrowing it down to something you, personally, can measure. A refined version of this question might be: "Does implementing a cloud based voting system improve the speed of decision making in a small company in Coventry?" This refined question is implementable: You can now identify a small company to work with, document their current decision making processes, implement a cloud based voting system, compare decision making speeds over a limited time period (say 1 month) and evaluate your findings. *A small piece of genuinely new knowledge is produced.*

"Can data analysis of electricity usage in the home be effective in identifying circumstances in which elderly or disabled residents may require aid"

1.2 Keywords

Help: Include up to 6 keywords separated by a semi-colon; what keywords are appropriate to describe your project in an online database like Google Scholar? Keywords should include the general research area and the specific technologies you will be working with. *Example.* A project that proposes a novel way of visualising large amounts of twitter feed data may have the keywords: Data visualisation; twitter; hashtags; database design; graphics libraries. For further help, take a look at the ACM keywords list http://www.computer.org/portal/web/publications/acmtaxonomy

Electricity Usage Trends; Clustering; Data mining; Data Analysis; Predictive algorithms; Household

1.3 Project title

Help: The project title is a statement based on your detailed research question. For example, the research question *'to what extent does a mobile application reduce the number of errors made in class registers at Coventry University in comparison to current paper based registers'* may be stated in the project title*: "A Wi-Fi driven mobile application for large group registers using iBeacons".*

"The use of data analytics to identify circumstances in which elderly or disabled residents may require aid through changes in electricity usage patterns."

1.4 Client, Audience and Motivation:

Help: Why is this project important? To whom is this project important? A research project must address a research question that generates a small piece of new knowledge. This new knowledge must be important to a named group or to a specific client (such as a company, an academic audience, policy makers, people with disabilities) to make it worthwhile carrying out. This is the ***motivation*** for your project. In this section you should address who will benefit from your findings and how they will benefit. Example: If you intend to demonstrate that a mobile application that automates class registers at Coventry University will be more efficient than paper based registers - the group who would be interested in knowing/applying these findings would be both academic and administrative staff at Coventry University and they would benefit by time saved and a reduction in their administrative workload. If you are making a business case for an organization explain how the organisation will benefit from your findings.

In a research survey carried out by the NHS it was revealed that around one in three adults over the age of 65 are susceptible to fall over at least once a year (NHS, 2015). Around half of whom are likely to fall more frequently. Another research concluded that two thirds of those who fall again in the next 6 months (Senior Health & Wellness Blog, 2018). In the UK alone falling being the cause of death is the most common unfortunate occurrence amongst seniors aged 75 and over (Age UK, 2017).

The application of this research project will be for the elderly and disabled people who reside on their own. Problems like getting around the house and doing daily chores etc can become a struggle at their fragile age and accidents often happen. As mentioned in the stats above, they can fall over or hurt themselves and with no one around, this can escalate into a grave matter.

Many steps have been taken to prevent such events from happening but most of the times they require additional expensive equipment like infrared or heat sensors in homes to monitor movement of the residents. However, electricity usage data is readily available for most households and the required data is being recorded but going to waste (the implementation of smart meters will later ensure the recording and storage of the consumption data for every household). Trends and patterns from this data can be a less expensive alternative solution to the same problem.

Electricity consumption data collected over a specific period can be analysed to derive trends which can then be compared to everyday usage to categorise the data as either normal or abnormal.
Abnormalities being anything like an electricity spike not being shown roughly at around 7pm when the inhabitant uses the kettle for their evening tea.

Cost of the physical persons going around
When abnormalities as such are visible, the system can choose to notify a user of the unusual occurrence in the data pattern.

In simple terms, this project aspires to implement a system that is capable of noticing abnormalities and deviations from normal trends in electricity consumption entries.

1.5 Primary Research Plan

Help: This is the plan as to how you will go about answering your detailed research question - It must include a primary research method (an extended literature review is not an acceptable primary method). Think and plan logically. Primary

methods may include experiments, applications or software demonstrators, process models, surveys, analysis of generated data …

Example: In the class register example above "to what extent does a mobile application reduce the number of errors made in class registers at Coventry University in comparison to current paper based registers" - the research plan may involve: 1) Collecting and analysing paper based registers in a given class on five occasions. 2) Identifying the error rate average on these occasions 3) Designing and implementing a mobile application that automatically records attendance in class. 4) Deploying the application in the class on five occasions. 5) Identifying the error rate average of the mobile application on these occasions. 6) Comparison of data and summary of findings.

This research will use iterative Agile techniques as its basic execution plan. Sprints every 2 weeks or so will be allotted to minimum functionality at first which will then be refined and further worked on every consecutive sprint.

1. Identify and choose an electricity consumption data set.
2. Identifying data analysis methodology such as clustering, neural networks etc.
3. Perform data analytics on the data set to extract trends and patterns
4. Refine results though false positives and false negatives
5. Implementation of data learning algorithm and software to notify user or system upon detection of an abnormal instance of data.
6. Apply statistical analysis on the results to identify the reliability of the system
7. Optional (depending on time): Conduct a survey on the willingness of people for the implementation of the system

This is the end of section one.

SECTION TWO: ABSTRACT AND LITERATURE REVIEW

2.1 Abstract

Help: An abstract is a short summary of a research project that enables other researchers to know if your report or research paper is relevant to them without reading the whole report. It is usually written retrospectively so that it can include findings and results. It is fully expected that you will rewrite your abstract when you come to write your final paper. For now, you should write an abstract of about 250 words that define the project described in section one. Before writing your abstract you MUST read some abstracts from conference or journal papers on *Google Scholar* or from *portal.acm.org* (to understand their style) and then provide your own abstract that outlines what your question is and what you 'did' to answer it.

The worldwide widespread of smart meters means more understandable & workable data will be produced for every specific household. In the UK the smart meter bill was passed on the 28th of November 2017 and the implementation of which will lead to raw electricity consumption data being produced and hence used to measure usage more efficiently (Gov.uk, 2013). This data can be mined and analyzed for more than just producing efficient bills and smarter readings. This project intends to provide an achievable, affordable and a minimal change or resource implementation required solution for people by analyzing their electricity data and extracting specific trends and electricity consumption footprints. Every person has different routines and lifestyles which impact and reflect a different electricity usage trend. Learning those trends can be useful to make a judgement or sense if something is unusual.

This will be done by using data mining techniques such as Clustering and regression. These techniques will be applied on pre-existing open source electricity usage data sets, the results from which will be used to investigate if a deviation from norm has been experienced. Where norm is a customized trend for every household. As an extension to that, the program will choose to notify someone if any unusual electricity consumption behavior occurs that is a deviation enough from norm to be considered suspicious.

2.2 Initial/Mini Literature Review (500 words – 750 words)

Help: A literature review is a select analysis of current existing research which is relevant to your topic, showing how it relates to your investigation. It explains and justifies how your investigation may help answer some of the questions or gaps in this area of research. A literature review is not a straightforward summary of everything you have read on the topic and it is not a chronological description of what was discovered in your field. Use your literature review to:

- compare and contrast different authors' views on an issue
- criticise aspects of methodology, note areas in which authors are in disagreement
- highlight exemplary studies
- highlight gaps in research
- show how your study relates to previous studies

There is plethora of research work that contributes to the area of computer science that deals with data mining and analysis, but I could find only a handful of work that directly associates my research. The literature that I chose adopts different methods of data analysis to interpret and analyse the fitting of their work in a specific application.

A research extract from a book analysed electricity data of a household to extract its specific trends which were then, along with the number of inhabitants of the household, used as input variables to a fuzzy model, to predict how probable it was for a house apparatus to be used or started within the next 60 seconds (Zhu, 2016). The electricity data set collection for this research was long term and collected over a period of 12 months. The research relates to the “Electricity data analysis” side of my research. A case-study on fuzzy tools was presented along with neural network based algorithm was presented to forecast the home electricity consumption 24 hours ahead of time.

Researches relating to Anomaly detection hugely relate to my project. The second part of my research is to detect, effectively, occurrences and instances of data that can be considered as anomalous. A survey-based journal article spoke in depth about the different kinds of outliers in data analysis and the challenges faced in isolating a given anomaly from normal regions (Chandola, Banerjee and Kumar, 2009). The research will particularly prove helpful in narrowing down concepts and techniques most effective in successful outlier isolation. In addition to that, it also spoke about techniques for Anomaly detection such as Neural networks, rule based, Bayesian Network, Support vector, Clustering. Nearest neighbour based techniques etc.
Another research relating to anomaly detection was more practical and aimed to detect abnormalities in 2 transformers of the UK based on data and variables such as temperature, vibration, moisture, load current etc between an aged and a new transformer (Catterson, McArthur and Moss, 2010). This paper concluded that the Conditional Anomaly Detection (CAD) technique can be used in the online monitoring of the transformers.

A research specifically relating to electricity and anomaly detection with extensive results in the form of figures, diagrams and tables carries out data analysis on electricity consumption data using 2 experimenting approaches namely statistical approach and clustering (Jakkula and Cook,

2010). It concluded that the wrong identification of an outlier ratio in Clustering:Statistical approach was 1:43. Meaning, in this given instance, clustering would be inaccurate 2.3% of the time whereas Statistical approach would wrongly identify an outlier 97% of the time.

Existing work done related to my research work takes things as far as analysing electricity usage data to either extract trends or test different data analytical techniques to state which one works better. My project intends to take it a step further, by using that analysed data and link it to an algorithm for notification purposes. The Anomaly Survey research discussed entirely different application areas of anomaly detection but none of them came close to the intended work and application of this project which makes me confident that the deliverable is new and hence will contribute to a novel application of outlier detection.
In addition to that, the anomaly detection survey spoke of the many techniques available for the data analysis and the results of the Electricity anomaly detection research stands to show that clustering is a viable data analysis technique for electricity consumption data.

2.3 Bibliography (key texts for your literature review)

Help: Please provide references, in correct Harvard style, for at least three key texts that have informed your literature review. If you are implementing an application, select texts which demonstrate how other researchers have tackled similar implementations? The references should be recent and sufficiently technical or academic. Your markers will be looking for you to identify technical reports, conference papers, journal papers, and recent text books. Avoid *Wikipedia* entries, newspaper reports that do not cite sources, and general or introductory texts.

Age UK (2017) *Briefing: Health and Care of Older People in England 2017*, UK: Understanding Society.

Catterson, V., McArthur, S. and Moss, G. (2010). Online Conditional Anomaly Detection in Multivariate Data for Transformer Monitoring. *IEEE Transactions on Power Delivery*, 25(4), pp.2556-2564.

Chandola, V., Banerjee, A. and Kumar, V. (2009). Anomaly detection. *ACM Computing Surveys*, 41(3), pp.1-58.

Gov.uk. (2013). *Smart meters: a guide - GOV.UK*. [online] Available at: https://www.gov.uk/guidance/smart-meters-how-they-work [Accessed 9 Feb. 2018].

Jakkula, V. and Cook, D. (2010). Outlier Detection in Smart Environment Structured Power Datasets. *2010 Sixth International Conference on Intelligent Environments*.

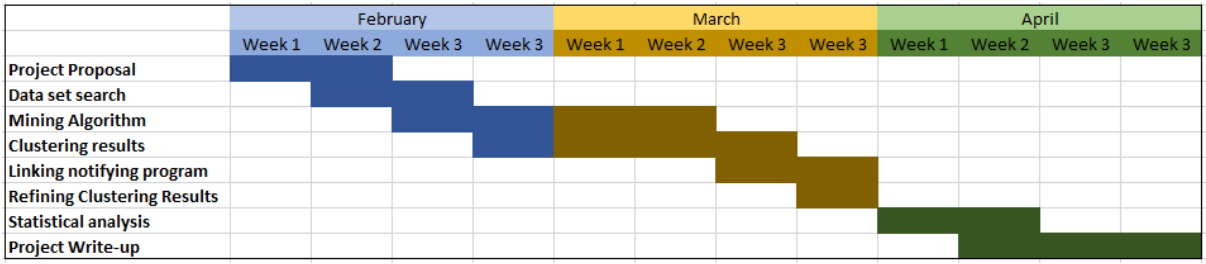
NHS, 2015. *Falls*, NHS. Available at: https://www.nhs.uk/conditions/falls/ [Accessed February 9, 2018].

Senior Health & Wellness Blog. (2018). [Blog] *10 Statistics About Elderly Falls*. Available at: http://shellpoint.org/blog/2012/08/13/10-shocking-statistics-about-elderly-falls/ [Accessed 3 Feb. 2018].

Zhu, Q. (2016). *Complex system modelling and control through intelligent soft computations*. [S.l.]: Springer International Pu, pp.437-467.

THIS IS THE END OF SECTION TWO

Time Management Plan – Gantt Chart



*The above is pre-mediated forecast plan for the breakdown and effective execution of the research project.

DETAILED PROJECT PROPOSAL GRADING FORM

The grade sheets for marking the 300COM / 303COM Detailed project proposal are attached on the next page.

Grading Notes:

The proposal is marked out of 20 divided into 10 marks for the quality***, achievability and level of challenge demonstrated by the student's research question and proposed primary method*** **of solution generation** and 10 marks for the ***thoroughness of the proposa***l.

Modal grading: In awarding marks please consider the following modal template:

	Research question and primary research method in relation to learning outcomes	**Thoroughness of the proposal.**
>70%	A well-considered project proposal that fully satisfies the Learning outcomes for which there is a succinct and focused aim with an associated project A question or hypothesis that is well above norm for final-year undergraduate project level (approaching Masters level for >80%); The project involves improving or developing a complex programme, tool, application or the enhancement of a theory or methodology or their application in a new context. The project demonstrates a high degree of innovation and creativity	All fields completed demonstrating a clear blueprint for the research process and includes the necessary information with respect to the research question. Research methods are well-considered with clear reasoning for choice of those methods over others; A clear justification of the need for the project in relation to client or audience. Projects proposals involving 'business case' reports clearly identify the organisation involved and consider how the case will be evaluated. A sound grasp of the means of evidence by which the conduct and management of the project may be judged.
Threshold (40%)	A proposal that identifies an activity with some consideration of a broader context. A research question which lacks enough substance, context and scope to allow for depth of analysis, but which is marginally acceptable against a threshold for final year undergraduate projects; A primary method(s) which only just relates to the production of an appropriate solution to the research question.	Completion of sections is cursory or minimal with some cohesiveness and contextualisation. Sections demonstrate some understanding of the research process involved which loosely links with idea outlined (key question, method, audience); Research methods are discussed but demonstrate little consideration as to whether they are the most appropriate and lack refinement and further detail. Identification of some methods of evidence for conduct and management of the project but unclear thinking about planning for reflection or accounting for conduct.