**Final 3rd year Project Proposal**

|  |  |
| --- | --- |
| Project title | Analysis and optimisation of building management systems through data analytics and interpretable machine learning |
| Student name | David Sal |
| Supervisors | Ryan Grammenos |
| Date | 20/10/2021 |

# Brief project description

Using data provided by an industrial partner I will analyse and try to predict the energy expenditure of different buildings. The data consists of measurements of the energy usage taken every half an hour over 1-2 years which gives us tens of thousands of lines of data.

First, I will need to clean and transform the data for analysis which includes processes such as identifying and handling outliers in the data. Second, I will identify and collect data of factors that affect the energy usage of buildings such as environmental factors (i.e., outdoor temperature, humidity etc.). This will be followed by the analysis of the data where I will search for data patterns and find similarities between different sets in order to put them into clusters of buildings with similar energy usage behaviour.

Once the data analysis process is finished a model can be built for predicting future events. This model needs to be trained to maximize the accuracy of predictions. To make the model applicable for all buildings outside the scope of our data it should be able to learn the energy profile of new buildings autonomously and be able to make real time predictions with a high degree of accuracy.

# Goals and objectives

List expected outcomes and the concrete steps to achieve those. (Bullet points with a maximum of three goals and 2‒3 objectives per goal.)

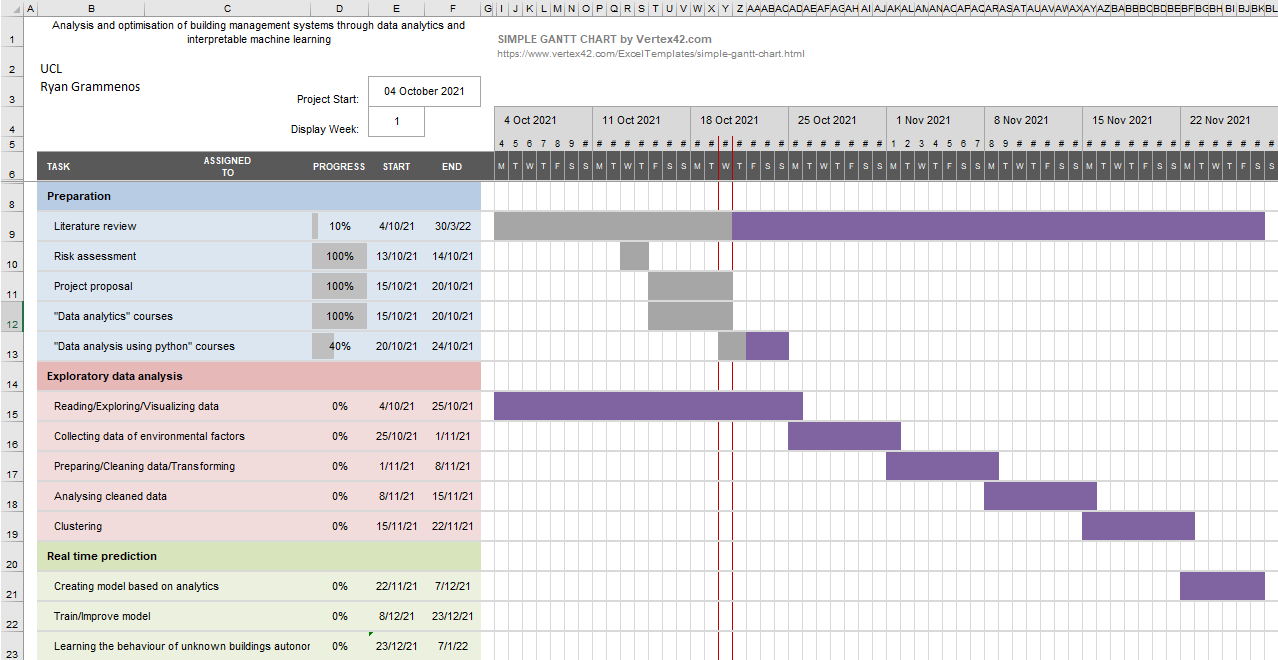
1. To perform data analysis on the dataset
   1. To clean/transform the data properly
   2. Collecting relevant data on environmental factors
   3. Analysing data and putting buildings with similar energy profiles into clusters
2. To make predictions of future events with a high degree of accuracy
   1. To make accurate predictions for the given buildings inside the scope of the datasets
   2. To make accurate predictions for new buildings outside the scope of the datasets

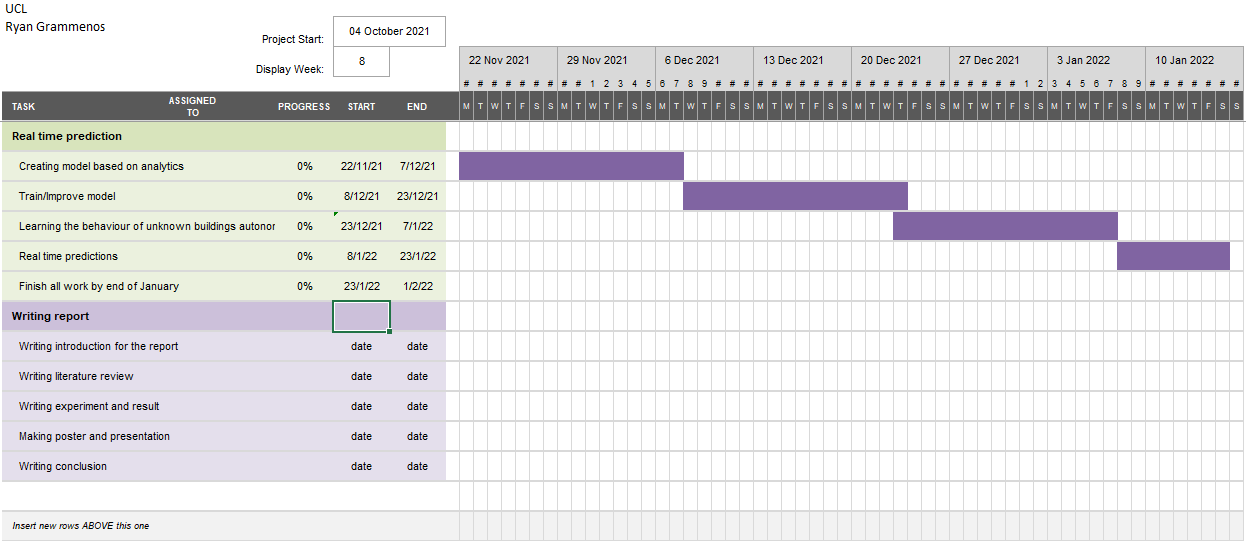
# Bibliography

List relevant literature and state in one sentence each why it is relevant. (Up to ten references. Use IEEE formatting.)

1. K. Yu, W. Shi, N. Santoro and X. Ma, "Real-time Outlier Detection Over Streaming Data," *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2019, pp. 125-132, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00063.  
   This paper analyses multiple outlier detection methods which can be considered for my project.
2. P. D. Yoo, M. H. Kim and T. Jan, "Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation," International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), 2005, pp. 835-841, doi: 10.1109/CIMCA.2005.1631572.  
   This paper gives an overview of time series prediction methods and how to incorporate additional information into the predictions. The methods and models are for the stock market, but I believe they can be used for my project since I am dealing with time series data which is influenced by environmental factors.

# Project schedule





# Failure risks

The project could fail if any part of the data analysis process is not conducted properly. For example, failing to identify important outliers leads to wrong energy profiles which makes any progress afterwards wrong or impossible to do. Transforming the data is also an important step as the data needs to be comparable to group the energy profiles into clusters.

A further failure risk is regarding the process machine learning to make real time predictions. It is possible to overfit the model which would make future predictions inaccurate. It would also make learning new energy profiles autonomously impossible.

Both these risks can be avoided by making sure all steps are conducted properly with the necessary care and precision.

# Safety risks

Table

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Activity** | **Hazard** | **Control** | **Risk level with control measure in place Severity x Likelihood** |
| **Long term computer usage** | Incorrect posture when sitting in front of a computer  Incorrect chair height  Incorrect lighting in the room  Radiating electromagnetic waves from the VDU  Sedentary work for a long period of time | Reduce the risks by ensuring correct sitting posture  Reduce risk by adjusting chair height  Reduce risks by eliminating light sources that cause glare/reflections  Taking regular breaks, resting eyes and using blue light filters  Taking regular breaks and exercise | Low  Low  Low  Low  Low |
| **Maximizing productivity and maintaining good mental health** | Temperature, humidity and oxygen in the room  Noisy environment  Individual work  Late night working with early mornings | The room should be well-ventilated with an acceptable room temperature.  Making sure to work in a quiet study space or using headphones with noise cancellation.  Finding a balance between work and social life.  Ensuring to have at least 7 hours of sleep every day by creating a daily schedule or setting reminders to go to sleep. | Low  Low  Low  Low |