******

**College of Business, Technology and Engineering**

Department of Computing

Project (Technical Computing)

[55-604708]

2020/21

|  |  |
| --- | --- |
| **Author:** | **Abubakar Abdalla** |
| **Student ID:** | **29040811** |
| **Year Submitted:** | **2021** |
| **Supervisor:** | **Chris Roast** |
| **Second Marker:** |  |
| **Degree Course:** | **Computer Science** |
| **Title of Project:** | **Gas reader prototype** |

|  |
| --- |
| **Confidentiality Required?**  **NO** |
|
|

|  |
| --- |
| I give permission to make my project report, video and deliverable accessible to staff and students on the Project (Technical Computing) module at Sheffield Hallam University.  **YES** |
|
|

Contents

[1 Introduction 3](#_Toc65263665)

[1.1 Project background and goals 3](#_Toc65263666)

[1.2 Aim and objectives 3](#_Toc65263667)

[2 Investigation 4](#_Toc65263668)

[2.1 Machine learning algorithms 4](#_Toc65263669)

[2.2 Development tools 4](#_Toc65263670)

[2.3 Conclusion of research 4](#_Toc65263671)

[3 Design 5](#_Toc65263672)

[3.1 Platform and library 5](#_Toc65263673)

[3.2 5](#_Toc65263674)

[4 Development 6](#_Toc65263675)

[4.1 6](#_Toc65263676)

[4.2 Tools 6](#_Toc65263677)

[4.2.1 Trello 6](#_Toc65263678)

[ How trello was used to develop the algorithm 6](#_Toc65263679)

[4.2.2 Sites 6](#_Toc65263680)

[5 Testing 7](#_Toc65263681)

[5.1 Types of testing methods taken into consideration to utilise 7](#_Toc65263682)

[5.2 Chosen testing method 7](#_Toc65263683)

[6 Evaluation 7](#_Toc65263684)

[6.1 Conclusion 7](#_Toc65263685)

[7 Bibliography 7](#_Toc65263686)

Acknowledgement

I would like to thank…

Table of figures

[Figure 1.2.1 what data scientists spend most time doing. 5](#_Toc67334226)

[Figure 2.2 Decision Tree(CART) 6](#_Toc67334227)

# Introduction

## Project background and goals

Gas bills tend to be sky high and sometimes inaccurate resulting to excessive payments to their gas bills. According to look after my bills (UK’s average gas and electricity bill, 2021) gas prices have rocketed over the last 15 years along with electricity. In 2004, the average home paid just £277 for gas and £288 for electricity, a total of £565. Fast-forward a decade and this final figure more than doubled to £1,344. We’ve been on a pretty high yearly incline that represents a 140% rise in bills.

This project will aim to develop a prototype that will be provide rich data information of the household gas bills’, this would help household members better their apprehension on the utilisation of their household’s gas usage. The solution will be a machine learning algorithm that learns the behaviour of a household’s gas usage. This will be done by household appliances gas usage data being validated against the algorithm allowing the data to be organised into sets that best describes it. This would help the algorithm provide richer information that would help the user understand where the excessive gas payments come from.

## Aim and Objectives

The aim of the project is to create a protype algorithm that will provide rich data information on gas usage. By collecting data from smart meters and creating an algorithm that will help learn the behaviour of the boiler will allow the prototype to know when an appliance is on or off as well as how much gas is an appliance using at a chosen time. A list of objectives for the project are as follows:

1. Explore project feasibility investigate viable.
2. Research project planning tools.
3. Create and submit project specification and sign ethics form.
4. Setup version control and project planning tools.
5. Research machine learning methods.
6. Research online data for smart home gas readers.
7. Practise machine learning data with sample data.
8. Practise implementing chosen machine learning methods for prototype data.
9. Complete machine learning algorithm.
10. Test protype for decision tree algorithm.
11. Make improvements for prototype.
12. Evaluate project.
13. Complete report for project.

# Investigation

Research is conducted to help better an understanding of the environment surrounding the problem, this is to allow a broader view of the possible solutions benefiting the conclusion of the investigation to be a solution befitting the problem. In this section Investigation will be composed of different elements of the problem as well as the chosen methodology. These elements consist of different machine learning algorithms, development tools, an outline of smart meters within households.

## Machine learning algorithms

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals which is to learn from experience. Machine learning algorithms use computational methods in order to learn information directly from data without having to rely on fixed equation as a model. The algorithm improves its performance through adaptation as the number of examples for learning increases.

### Feature programming

Feature engineering is an engineering practise used on algorithms to learn data and be able to make the appropriate decisions when organising them. This approach consists of 2 goals, preparing the proper input dataset, compatible with the machine learning algorithm requirements and improving the performance of machine learning algorithms. This engineering practise is unique to others as it uses other methodologies and improves the performance to further utilise in its own algorithm. However, this approach tends to be time consuming as a lot of time is spent during the preparation stage. According to towardsdatascience(Fundamental Techniques of Feature Engineering for Machine Learning, 2021) Forbes surveys resulted to data scientists spending over 80% of their time on data preparation.

Chart, sunburst chart

Description automatically generated

Figure 1.2.1 what data scientists spend most time doing.

### Decision tree algorithm

A decision tree algorithm machine learning algorithm that is used to mine and analyse data in order to find a strategy that will best work for a dataset so that the goal may be achieved. It uses decision in a form of a tree like model to permit the algorithm to sort out the data accordingly. A decision tree algorithm starts with a root and works its way down breaking data into questions that will best help categorise the data into the right sets that best describes its features. This algorithm helps developers understand the data they are working with more as multiple functions and classes are implemented solely to better the utilisation of the data within the nodes and decision sets.

There are many types of decision tree algorithms, such as; ID3, C4.5, CART CHAID and MARS. One that I focussed on is CART which is a classification and regression tree algorithm this means that it focuses on binary splits to sort out data whereas the others create multiple branches of a single parent node making it more difficult to interpret and work with.

The Advantages of using CART to analyse data is the simple fact that it is simple to understand as its methods are simple and goes on the bases of questions that will lead have yes or no answers. This allows it to easily be interpreted and visualise making it efficient to implement as errors within the algorithm would be easily identifiable. Furthermore, not only can it handle both numerical and categorical data, but also it requires less effort from users to prepare. This can be useful with finding relevant data that will help make the difference with understanding household gas bills.

Contrarily, CART can be unstable due to small variations of the data which would cause a completely different decision tree to be generated creating a variance. This would make interpreting the data into the classes somewhat difficult due to the assurance you will be required to have with how the data will be interpreted. In addition, although CART is a simple decision tree algorithm, users can tend to create over complex trees causing the data to not generalise very well making it overfit.

Diagram, timeline

Description automatically generated

As shown in figure 2.2 the example is based on the weather. As simple it is the features importance are clearly visible making it easily to interpret and understand.

Figure 2.2 Decision Tree(CART)

## Development tools

### Programming languages

### Github

### Trello

* List of development tools that can be used
* Analyse all of them by its categories weighing its advantages and disadvantages
* State chosen tools to be used and why.
* Different programming languages that each development tool supports as well as most effective programming language.

## Conclusion of research

* Which features and tools you will use and which you wont explaining why

# Development

## 

## Tools

* How the tools that were used to help develop the algorithm

### Trello

* How Trello was used to develop the algorithm

### Jupiter notebook

* How Jupiter Notebook was used to develop the algorithm

# Testing

## Types of testing methods taken into consideration to utilise

## Chosen testing method

# Evaluation

## Conclusion

# Bibliography

(UK’s average gas and electricity bill - Look After My Bills, 2021)

Look After My Bills. 2021. UK’s average gas and electricity bill - Look After My Bills. [online] Available at: <https://lookaftermybills.com/blog/average-gas-electric-bill-uk/> [Accessed 22 March 2021].

(Press, 2021)

Press, G. (2021). Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. Forbes. Retrieved 22 March 2021, from <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=5a700ea96f63>.

(Castañón, 2021)

Castañón, J. (2021). Feature Engineering: What Powers Machine Learning. Medium. Retrieved 22 March 2021, from <https://towardsdatascience.com/feature-engineering-what-powers-machine-learning-93ab191bcc2d>.

("What Is Machine Learning? | How It Works, Techniques & Applications", 2021)

What Is Machine Learning? | How It Works, Techniques & Applications. Uk.mathworks.com. (2021). Retrieved 22 March 2021, from <https://uk.mathworks.com/discovery/machine-learning.html>.

(Ali, 2021)

Ali, A. (2021). Decision Tree with Practical Implementation. Medium. Retrieved 22 March 2021, from <https://medium.com/machine-learning-researcher/decision-tree-algorithm-in-machine-learning-248fb7de819e>.