An Artificial Intelligence Approach to Modelling Bank Customer Energy Efficiency

Oisín Brannock (20235671)
School of Computer Science
National University of Ireland Galway

Supervisors
Dr. Karl Mason

In partial fulfillment of the requirements for the degree of
MSc in Computer Science (Artificial Intelligence)

11th July 2022

ACADEMIC DECLARATION

I, Oisín Brannock, do hereby declare that this thesis entitled "Artificial Intelligence & Bank Customer Energy Saving" is a bona fide record of research work done by me for the award of MSc in Computer Science (Artificial Intelligence) from National University of Ireland, Galway. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signature: ______________________________

BANK X DECLARATION

I, Oisín Brannock, do hereby declare I adhere to and accept the principles for using data from Bank X as part of my research thesis. No data has been shared outside of the network of the bank. No personally identifiable information has been released. All info provided in this thesis is at an aggregated level. This information does not form an opinion of Bank X, and is purely used for academic research purposes.

Signature: ______________________________

Acknowledgements

I want to thank Bank X for allowing me the use of their data to pursue this research.

I want to thank Dr. Karl Mason, who has been a strong pillar to lean on through any ambiguity I had throughout the entire research project.

Abstract

The abstract should summarize the substantive results of the work and not merely list topics to be discussed. An abstract is an outline/brief summary

It should be terse and usually written in the present tense and impersonal style: “A new graph community detection algorithm is proposed based on

Keywords: keyword1, keyword2, keyword3, keyword4, keyword5

Contents

List of Figures

List of Tables

List of Acronyms

AI Artifical Intelligence. 4

ANN Artifical Neural Network. 17, 19, 29

BER Building Energy Rating. 25

CART Classification and Regression Tree. 12

CNN Convolutional Neural Network. 29

CRISP-DM Cross-Industry Standard Process for Data-Mining. iv, 5–7

ELM Extreme Learning Networks. 29

FLNN Functional Link Neural Networks). 29

GA Genetic Algorithm. 29

IEEE Institute of Electrical and Electronics Engineers. 23, 24

KNN K-Nearest Neighbours. 2, 8, 9, 12, 22

LSSVM Least Squares Support Vector Machine. 28

LSTM Long Short Term Memory. 21

Chapter 1

Introduction

1.1 Motivation

Artificial Intelligence can help model energy usage for customers of Bank X and help reduce energy waste.

The topic of energy efficiency is one of the most prevalent topics in our society today. The move towards environmentally friendly practices has seen a dramatic rise in popularity, in both the private and public sectors, as well as within the general public. A variety of approaches have been taken in this regard. Journals published as early as the 1980s have broached different approaches in search of the ideal model [2]. The scope of this topic governs every industry in the world and is imperative to a sustainable future for humankind. Initially, the studies in this field focused on efforts to improve the efficiency of buildings during their construction period [3]. This has evolved in the age of information into using computational models to predict what can impact energy usage in a building, whether it be a school, office or water treatment plant [4; 5; 6; 7].

The goal of this thesis is to model energy usage for customers of Bank X with their providers using spend data as opposed to raw energy data, and to use this model to show areas where energy savings could be made. Customers would gain an awareness of how much they spend compared to how much they only **need** to spend. This analysis can open doors for new green focused products that help Bank X as a brand, help customers save money, and most importantly help raise awareness and inspire action for a sustainable environment. However, this raises questions in regard to which approach is the most optimal. This thesis seeks to determine which method is the most viable and apply a new approach. Methods such as neural networks, decision trees, KNNs have been shown to be effective in this space using sensor data.

1.2 Thesis Structure

Chapter 1 outlines the motivation for this paper.
Chapter 2 focuses on the methods used to perform research and analysis of the data.
Chapter 3 is an in depth literature review of the work done in the space of energy and sustainability to get an idea of a baseline to work with.
Chapter 4 delves into the data used for analysis; where it comes from, how it is processed for use and what considerations had to be taken into account in its usage. The analysis portion deals with the modelling of the data; what models were chosen and why, how each is optimised for the dataset at hand, overfitting etc.
Chapter 5 highlights the experimental method from start to finish; steps on how analysis was done, settings of cross validation etc.
Chapter 6 examines the model results, and goes into depth on post analysis to

determine the significance of the results and what they mean

Chapter 7 provides an in depth discussion of the results presented in the Chapter 6; what significance they have in real world use cases? Are they as expected, or is there some flaw highlighted?

Chapter 8 presents a summary of the thesis, outlines the contributions of the research, provides answers to the research questions posed in Chapter 1, and finally discusses the implications and impact of this work.

All diagrams used in this thesis have been created by the author and are only referenced when required.

1.3 Research Questions

This research will answer the following research questions (RQs):

RQ_1 - What are the data requirements for modelling energy usage in Bank X customers?

RQ_2 - How should machine learning be applied to best predict customer energy efficiency?

RQ_3 - How can users effectively interact with and gain insight about the customer base from this customer model?

Chapter 2

Background

This chapter gives context to the methods used in this thesis for analysis. It also ties into how each one is relevant to the idea of energy saving.

2.1 Machine Learning

Machine learning is a branch of AI that allows machines to solve a vast range of problems faster, and more often than not, more accurately than a human can. Machine learning is concerned with designing programs that can learn rules and patterns from data, and adapt to new scenarios based on this training [8]. This allows machines to infer answers rather than having to be explicitly programmed. Many of the challenges we wish to overcome in today's world are not straightforward and cannot be simply programmed for a computer to solve in a binary manner. There are a plethora of techniques machines can use to solve problems, which are constantly being improved upon through research and analysis. Machine learning consists of 3 sub-categories: supervised learning, unsupervised learning and reinforcement learning. This thesis focuses on the use of supervised learning methods. In Fig. 2.1, the general flow of a machine learning process is outlined. It starts by first gathering, cleaning and splitting a dataset into training and testing datasets, and normalising/transforming data to suits the means of the task at hand. Next the model is trained on the training data. The model is then evaluated using statistical tests to check if the results are statistically significant and if they make sense in reference to the hypothesis. The model can then be deployed for use and improved using new incoming data. This is a very simple idea of how models are created. Section 2.2 expands on this for a more cohesive structure.

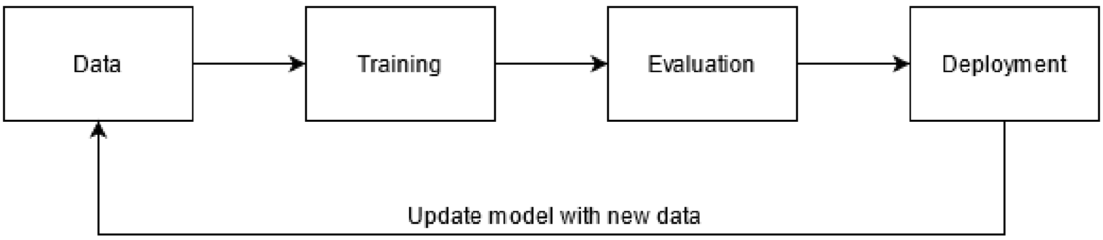

Figure 2.1: Simple Machine Learning Workflow

The availability of data today has led to a huge uptake of machine learning practitioners, expanding beyond the realms of scientific research into everyday life. This has included models that cover the likes of fraud detection in the financial industry, allow cars to be autonomous, or facial recognition that is used in phone and security technology [9; 10; 11]. The use cases are endless and constantly being expanded upon.

2.2 CRISP-DM

The Cross-Industry Standard Process for Data-Mining (CRISP-DM) is a methodology that seeks to standardise how models are developed and maintained across industries. This is to keep a consistent and thorough approach that captures all aspects of a problem [12; 13]. CRISP-DM is important in relation to the analysis of this thesis as it serves as a base layer of guidance. This thesis has been structured in a way that follows this format, ranging from the "business understanding" aspect in the first couple of chapters, all the way to modelling and deployment in the latter chapters. This is especially important here as the data used is customer data, and therefore the process should reflect this to the best possible standard in order to make the best possible impact for this cohort.

Business understanding involves developing a hypothesis for the problem at hand. Is this really a problem that could benefit from machine learning? Is the problem well defined? Is there data available to make a model? What tools and technologies will be used? What does success look like? Break down the project into phases, like chapters of a thesis.

Data understanding comes when the problem has been analysed and hypothesised accordingly. What kind of data do we have? Is it structured in the form of a dataframe? Does it have one row per observation? Is it in wide or long format [14]?

Data preparation involves making sure the data is in an appropriate format for the modelling technique to be used. Does the data need to be normalised? Does it need to be in wide or long format? Do we need to fill in null values? How do we fill in these nulls if they exist; with 0, with the mode or with the mean? Could we create a model to predict what these nulls may be? Could we engineer new features based on the ones we already have at out disposal?

Modelling involves choosing and creating models and determining the one that best suits the data. The best model at first may not be the best model after cross validation. If model performance is lower than desired, one may need to go back a step or two to determine if the data has been understood correctly or has been prepared for analysis correctly.

Evaluation takes the best model and determines if it meets the needs of the project to an acceptable threshold. How would one explain the findings to someone in the simplest terms? If they cannot do this, they do not fully grasp what they have done and must review the work thoroughly to ensure accuracy. The key question in this phase is: **Have we solved the problem we outlined in the initial phase of this project?** If the answer is yes then one can begin devising how to move forward with the desired model.

Deployment involves putting a model into a state where it can be accessed by people outside of the project to solve the problem at hand. How does one plan on achieving this deployment? What resources will be used? Who will maintain the model for incoming data? When these questions have been addressed, a final report can be outlined on the project and reviewed for where things could have been improved in hindsight. If deployment needs to be updated or improved in some capacity, the process can always be restarted.

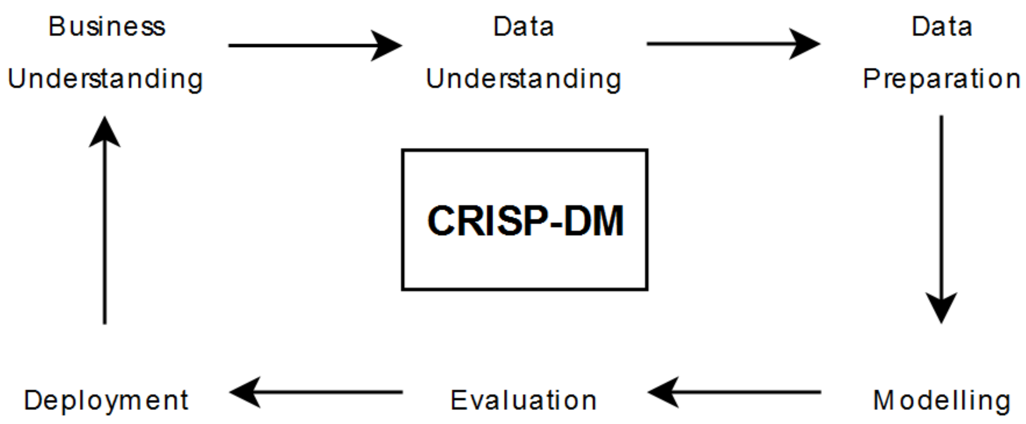

Figure 2.2: CRISP-DM Methodology

2.3 Supervised Learning

Supervised learning is a sub-category of machine learning. It involves the use of labelled data to train a computational model. In training the model, the machine is shown the correct outcome for each training example. The basis for this technique is that we can predict what will happen in similar future scenarios given what has already happened in past scenarios [15]. This assumes that the factors that initially led to these outcomes have not changed, which may not always be the case depending on the problem. Supervised learning can be split into classification and regression tasks. Classification deals with categorical, discrete and Boolean predictions. Regression deals with the estimation of continuous values. The focus in this thesis will be on regression as the goal is to predict energy spend, which is a continuous variable.

2.3.1 Regression

Regression involves the use of statistical methods to mathematically model the relationship between a dependent variable and one or more independent variables. The subsections below seek to explain regression techniques that were considered for the modelling of the data used in this thesis.

2.3.1.1 K-Nearest Neighbours

KNN works under a very simple premise; any datapoints that have similar characteristics to other datapoints in a dataset will have similar outcomes [16]. This is known as feature similarity. In KNN regression, the algorithm starts by calculating the distance of the new datapoint to each training datapoint. This is done either by calculating the Euclidian or Manhattan distance.

The Euclidian distance is calculated by taking the square root of the sum of the difference of squares between the new datapoint and training datapoint [Eq. 2.1].

The Manhattan distance is calculated by getting the shortest distance between the two vectors that represent the new data and training data, and get the sum of their absolute difference [Eq. 2.2].

The Minkowski distance is a generalisation of the Euclidian and Manhattan distance formulae. The formula relies on the constant p. If $p = 1$, then we have the Manhattan formula. If $p = 2$, we have the Euclidian formula [Eq. 2.3].

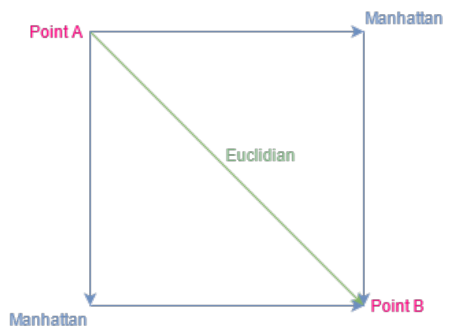

Figure 2.3: Distance Formulae

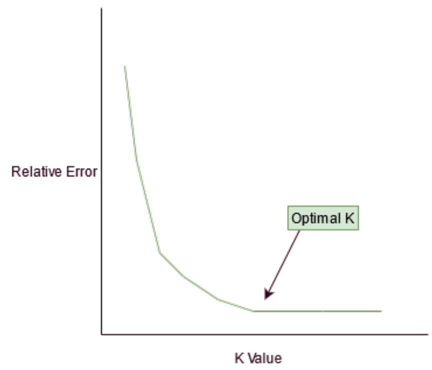

Figure 2.4: Example Elbow Plot

If a new datapoint is close to 2 or more training datapoints, the average of the training datapoints close by are taken to be the value of the new datapoint. This is subjective relative to the value chosen for k. The optimal k value can be found by modelling for a range of values of k, and plotting the relative error of calculations against the k values. This is known as an elbow plot (Fig 2.4). A rule of thumb used is $k = \sqrt{N}$, where N is the number of training datapoints. So if we had 400 training datapoints, $k = \sqrt{400} = 20$. An elbow plot may prove more useful than this rule however.

KNN is a useful modelling approach as it is simple and intuitive, and also does not assume anything about the data at hand. However, it does suffer from the curse of dimensionality, wherein the more independent variables that are used to predict a dependent variable, the number of dimensions increases [17]. This can be mitigated using PCA, which is a dimensionality reduction technique that can bring the number of dimensions back to 2.

$$D_{Euclidian} = \sqrt{\sum_{j=1}^{k}(x_j - y_j)^2} \tag{2.1}$$

$$D_{Manhattan} = \sum_{j=1}^{k} |x_j - y_j| \tag{2.2}$$

$$D_{Minkowski} = \left[\sum_{j=1}^{k}(|x_j - y_j|)^p\right]^{\frac{1}{p}} \tag{2.3}$$

2.3.1.2 Linear Regression

Linear Regression is a basic but powerful technique that allows for regression modelling. Simple linear regression involves one independent variable and one dependent variable. A linear relationship between the variables is assumed. A general example of a linear relationship that could be modelled using linear regression is someone's height and weight. Here the relationship is positively correlated, with taller people tending to weigh more.

Linear regression seeks to fit a line of best fit between the dependent and independent variable. This can be done using methods like gradient descent to reach a global minimum error cost function, or using a method of ordinary least squares, where one takes the smallest sum between squared errors, similar to Eq 2.1. The error is the difference between the predicted value and the actual value.

There will always be some sort of error in the line of best fit and therefore that alone is not a good enough metric to gauge how well the model performs. The R-squared metric can be used as follows:

$$R^2 = 1 - \frac{\sum_j (y_j - \hat{y_j})^2}{\sum_j (y_j - \bar{y_j})^2} \tag{2.4}$$

Where the top portion of the fraction represents the residual sum of squared errors i.e. the difference between the true value of the outcome and the predicted outcome. The bottom portion represents the **total** sum of squared errors, which is the difference between the true value of the outcome and the mean of the outcomes.

In the majority of real world problems, we don't have just one single variable that is dependent on an outcome. This gives rise to the use of multiple linear regression, where the formula expands to add extra terms.

The assumptions of linear regression limiting. The data must be fully understood before attempting to create a linear model. It assumes that there is a linear relationship between the independent and dependent variables. It also assumes that the independent variables are just that; independent in that they do not impact one another in any way and stand alone [18]. The formulae for simple and multiple linear regression are defined as:

$$y = b_0 + b_1 x_1 \tag{2.5}$$

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n \tag{2.6}$$

Where y is the dependent variable, b_0 is the y-intercept, b_n is the n^{th} slope and x_n is the n^{th} independent variable. Similar to KNN, the more dependent variables we add, the more dimensions we have to deal with [17].

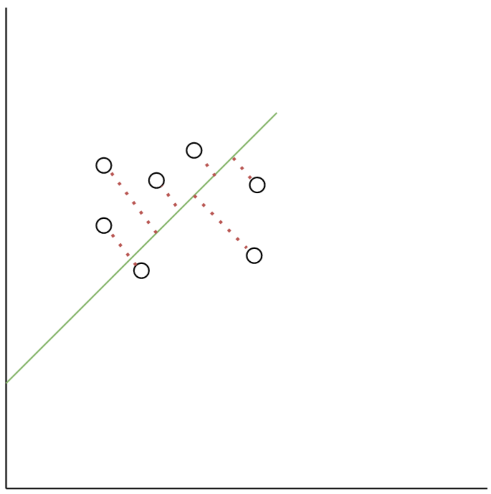

Figure 2.5: Linear Regression with relative error lines

2.3.1.3 Regression Trees

Decision trees are very useful models for regression analysis. Regression trees are developed generally using the CART methodology, which takes the independent variables and uses them to split nodes up to predict a dependent variable. A decision node is one that splits into another decision node and leaf nodes, or just leaf nodes. Fig. 2.6 shows a very basic example of a regression tree workflow to illustrate this. The alcohol content in a set of examples is being predicted using input variables such as nitrogen content, colour and alcohol by weight. Notice how as we move down the branches of the tree, the MSE values decrease further and further. This tree has also been pruned to have a maximum depth of 4 so as to not overfit the data.

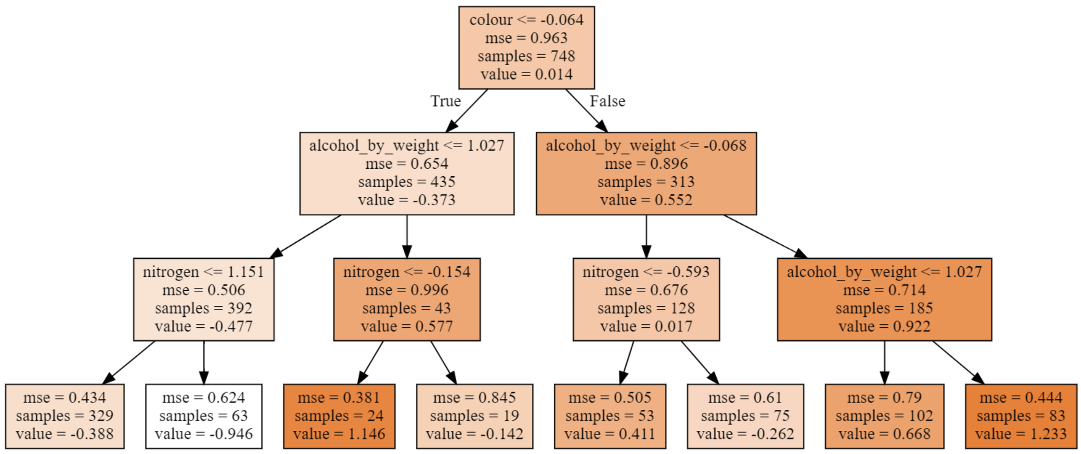

Figure 2.6: Regression Tree Flow [1]

The splits of decision nodes can be determined in regression using various methods.

One such method is called reduction of variance [19]. Variance refers to how sensitive a model is to fluctuations within a training dataset and is defined as:

$$V = \frac{\sum(X - \mu)}{n} \tag{2.7}$$

Where X is a specific datapoint, μ is the mean value of the data and n is the number of datapoints.

We want the child node with the least variance when splitting a parent node. This calculation is done for the dependent variables each time the parent nodes need to be split and chooses the child node with least variance to proceed with. The tree will look at each dependent variable and check to see which results in the lowest variance when split and choose this variable to split on and so forth until the tree ends. Regression trees need to be pre-pruned generally, which means we set a max depth for the tree so it doesn't overfit on the training data.

Mean Squared Error (MSE) is a metric used to gauge the accuracy/error of the model after training. This method is similar to reduction in variance (although it is not a regression algorithm inherently) in that each dependent variable is examined by the formula and the split that minimises the MSE is chosen for the parent node. MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2.8}$$

Where n is the number of data points, Y_i are the actual values and $\hat{Y}_i$ are the predicted values by the algorithm.

Regression trees are prone to overfitting on the training dataset, and though this can be avoided using cross validation, a good solution would be to increase the number of trees used and use the majority voting of these trees to make the best model. This is known as a random forest [20].

2.3.1.4 Random Forests

Random forests make use of ensemble methods to use many decision tree learners in order to enhance a models performance. Random forests allow an individual tree in the forest to grow very large without needing to be pruned, as we are not as concerned anymore about the high variance of a single tree. A method known as bootstrapping [21] is used to pick random samples from the dataset to train each tree with. This negates overfitting. While a single tree may overfit, as a collective the forest will not be biased to any specific training data.

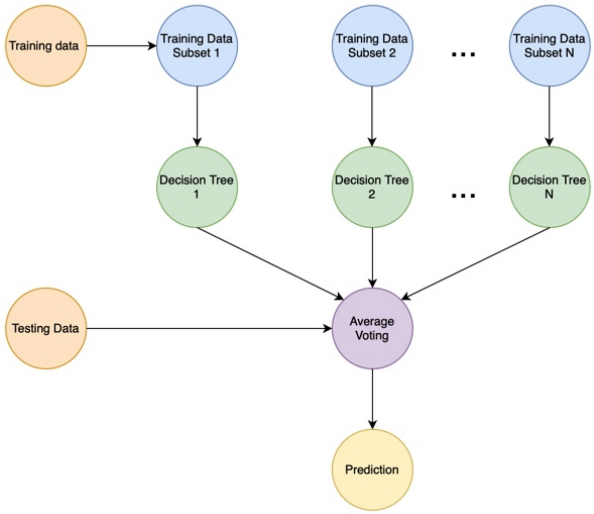

Figure 2.7: Random Forest Flow

When constructing each tree, the variables for each split are chosen randomly as opposed to using reduction of variance. This again is to ensure that each tree is unique and they do not correlate with one another to lead to a large bias. Hence, we have low variance.

Finally, we repeat this process for n trees. The average prediction of the trees is taken as the prediction for a specific test datapoint.

2.3.1.5 Support Vector Regression

SVMs make use of a hyperplane to best fit a dataset. In 2-D, this hyperplane would form a line. However in most scenarios, we would have a multitude of dependent variables and the dimensionality of the hyperplane would scale accordingly. This can then be transformed back to a 2-D view to show how the hyperplane looks. What makes SVR unique from other regression methods is that it incorporates a margin of acceptable error. This error lies between the chosen hyperplane and a boundary line mirrored on each side of the hyperplane. This is illustrated in Fig. 2.8.

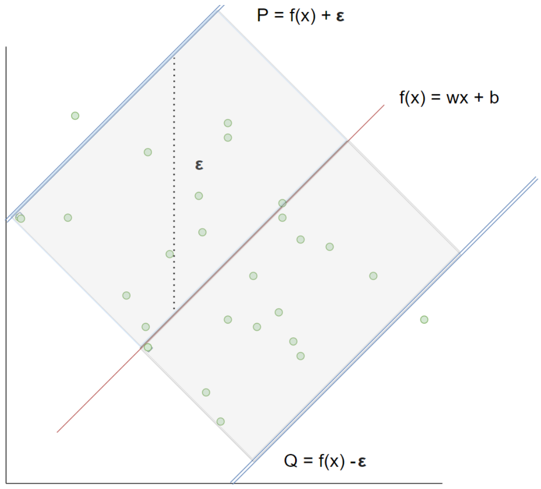

Figure 2.8: Support Vector Regression

The purpose of SVR is to determine a line of best fit, $f(x)$, that deviates at most a value of ϵ from other border line functions, P and Q. The grey shaded area in Fig. 2.8 shows the area of acceptance between P and Q. Any point outside of this area is outside of the margin of error of the model. A kernel function is picked by the algorithm in order to transform the data to higher dimensions to make it linearly separable [22]. Cross validation allows one to model a dataset using a range of kernel functions and choose the one that performs best on the given data over a chosen fold value. For example we may find a dataset is not linearly separable in 2-D and therefore need to use an RBF kernel [23] to transform it to higher dimensions where it can be separated using an n^{th} dimensional hyperplane, where n corresponds to the number of dependent variables used to predict an independent variable.

In general it is important to perform feature scaling on data for support vector machines in order to counteract dimensionality issues as much as possible. For example, if one dependent variable is orders of magnitude larger than another dependent variable, it would be best to normalise the data so that it all lies between 0 and 1 to allow the algorithm to be as low biased and low varianced as possible. This can be done as follows:

$$x_{normalised} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2.9}$$

SVR is robust to outliers in a dataset and can generalise to higher dimensions easily. In saying that, the training time can grow exponentially if a dataset is large, but this tradeoff is usually worth it for a good model.

2.4 Neural Networks

Neural Networks are algorithms that are modelled on the structure of the human brain, and try to replicate this process of information retrieval and deduction. They consist of nodes called neurons, which are simply nodes that data flows through. In it's most basic form, a neural network takes input data, puts it through neurons that perform mathematical processes on the input data, and finally return this new processed data as output data [24]. Neural networks have become very popular in recent years due to the availability of large quantities of data and increased computing power. Outlined below are two forms of neural network that have been selected as possible models for this thesis.

2.4.1 Artificial Neural Networks

ANNs are not a new concept in machine learning. A multitude of machine learning methods can be represented as neural networks. For example, simple linear regression can be viewed as 2 input neurons that are multiplied by a weight and bias (slope and y-intercept here) and added together to reach a single output layer as shown in Fig 2.9.

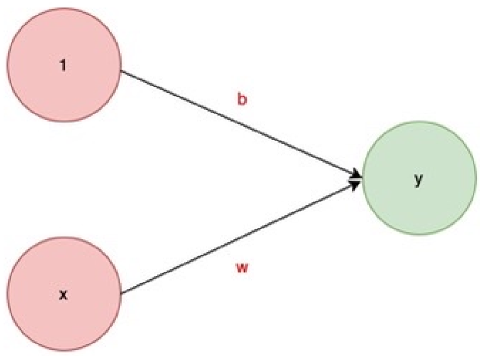

Figure 2.9: Linear Regression Network: $y = wx + b$

In more complex regression problems, we need hidden layers that perform processes on the input data before being output as a prediction. The hidden layers would make use of a linear or ReLU activation function [25] by convention, but any activation function can be used. Finally when the neural network outputs a signal from the output layer, this can be used to make a prediction. Metrics such as MSE can be used to measure the error of the prediction. This can be fed back to the network for backpropogation using gradient descent to update the weights and biases in order to improve the predictions [26; 27]. This is illustrated in Fig 2.10.

A ReLU activation function is more desirable than a linear one as we can get derivatives of the ReLU function in order to backpropagate the network to improve the results.

$$f(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ if } x \geqslant 0 \end{cases} \tag{2.10}$$

The ReLU function above has derivatives 0 and 1. Any negative value will end up with an output of 0. The function is used as it prevents the problem of vanishing gradients during model training. Vanishing gradients occurs when the gradients of the activation function get smaller as they reach a global minimum. This can get to the point where the gradients become exponentially small, and hence the gradient descent will never reach the global minimum. solves this issue but can also lead to some of the neurons giving an ouput of 0.

The use of a neural network as opposed to one of the earlier regression methods outlined could be chosen to avoid manual feature extraction that perhaps could be too complex, while also providing detailed modelling of the dataset that may not be as refined in a simple regression algorithm.

It's important to note that we can use more than one hidden layer. In fact, in more complex scenarios this is preferred as it will transform input data away from linear format the more layers that are used. One of the benefits of these deep neural networks is that they allow us to build more complex function approximations than simple methods like linear regression.

2.4.2 Recurrent Neural Networks

Recurrent Neural Networks are similar to ANNs but have one major advantage; they can remember previous input data after it has passed on, due to the incorporation of a memory like system. For example, time series forecasting could be very beneficial with a RNN [28]. Time series data is sequential and there-

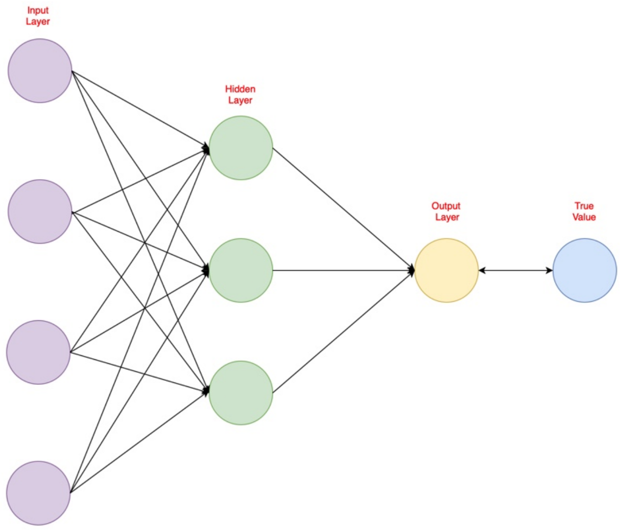

Figure 2.10: Feed Forward Neural Network

fore a normal feed forward neural network would not be suitable to model this relationship. We need context in order to be able to accurately determine the next timestep value. In time series analysis we also cannot assume all the input variables are independent from each other.

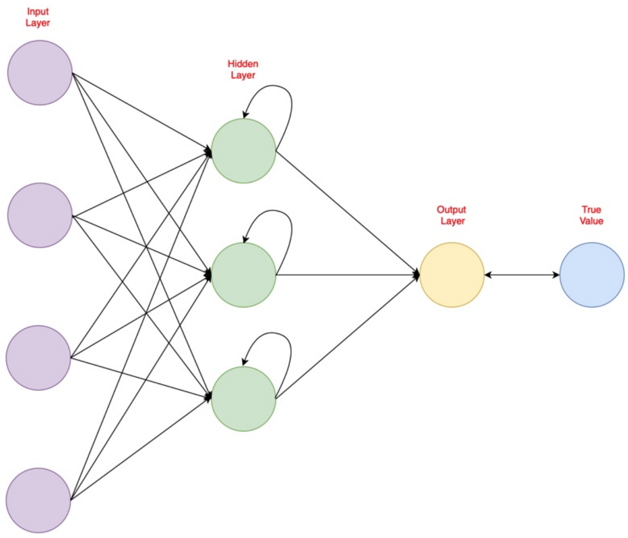

Figure 2.11: Recurrent Neural Network

Fig 2.11 shows the recurrent nature of the network in the hidden layer. When the hidden layer receives the data from the input layer, it also receives output from the previous step to use in training.

An issue that can arise with RNNs is that sometimes the context needed to understand a pattern can be distant in the network and as such, the RNN becomes unable to identify a pattern. Long Short Term Memory (LSTM) is a method used to negate this very issue. In a normal RNN, there is one layer in the a repeating hidden layer node. LSTMs are created so that they have the same structure, but have 4 layers that interact with each other within the same node [29]. LSTMs have been very successful in recent years and are used in many modern RNN models.

2.5 Energy Saving Applications

We have explored algorithms that have been considered in the analysis for this thesis, but what relevance have they to energy saving?

Chapter 3 will evaluate the various studies in the literature that apply machine learning to energy efficiency and energy saving. Sustainability is a very promising research area for machine learning to benefit from in the coming years, when it is needed more than ever.

For example, without doing any research we know that modelling for energy bills is a regression problem that can be addressed through any of the regression methods outlined previously. The dependent variable would be the energy cost, and the independent variables could be anything from time of year to appliance usage.

The neural network approach could be utilised if the best available data is in time series format using an RNN. We could also use a feed forward network with a ReLU activation function for normal tubular data for a simpler regression analysis if this data proves more useful and compare this to the regression methods such an KNN and random forest regression.

Chapter 3

Literature Review

3.1 Research Method

The search was conducted from Scopus within the NUIG Library system, IEEE Xplore and ScienceDirect, which comprised terms such as "Energy Cost Saving Artificial Intelligence" and "Home Energy Saving". Based on these searches, the most relevant titles were found and scanned for relevant abstracts. If an article was deemed relevant to this thesis' hypothesis, the papers cited by the article were examined to get even more background.

The second set of searches revolved around similar terms, input to Google Scholar. Older papers (pre-2010) were found and scanned for relevance. These papers lay the foundation for machine learning and AI techniques in the field of energy analysis and optimisation. While this thesis will revolve around customer spend data as opposed to direct energy readings, the papers found provide an excellent level of insight that helped guide this review.

Overall, the search proved a success and provided the basis for 3 research questions for this thesis:

RQ_1 - What are the data requirements for modelling energy usage for Bank X customers?

RQ_2 - How should machine learning be applied to best predict customer energy efficiency?

RQ_3 - How can users effectively interact with and gain insight about the customer base from this customer model?

These research questions serve as a foundation for the literature review.

Notes

The references included are based on the quality of publication, all of which are peer reviewed, as well as the quality of the abstract. Books were also analysed for relevance and new techniques (both books cited are new editions published in 2021). The older papers were given preference to get a baseline in the work done in the field of energy analytics and henceforth, move on to some of the more recent work, again based on title relevance and abstract quality.

The references provide a varied sample set to work with in terms of methodology and approaches. They are then scanned and analysed if they were deemed to have fallen in line with one or more of the research questions outlined. The references are in IEEE style.

3.2 Literary Review

Energy efficiency has been at the forefront of many industries across the world, particularly in the last 50 years due to the oil crisis of the 1970s. Both in the private and public sector, officials are looking for ways to incorporate methods of efficient energy for many reasons, but all avenues lead to one end result; a sustainable future for the human race by reducing our carbon footprint. Narciso et al. [30] presented an overview of 42 of the most reputable papers in the area of energy efficiency over the last 20 years, outlining the models chosen along with input variables, pre-processing techniques etc. It highlights that despite thorough research, there are still areas that are lacking in their real world application.

The field of energy efficiency is vast and, as such, not all of it is in scope for this review. The focus will be on the technical approaches, such as data processing and model choice. The literature review is broken down by the research questions that needed to be addressed. From there, the research is collated and examined as a whole in order to gauge what is most relevant to the scope of this project.

RQ_1 : **What are the data requirements for modelling energy usage for Bank X customers?**

Machine Learning in Buildings

Before devising any model plan, data needs to be the forefront of analysis. There may be a lot of data available to avail of, but is it all relevant? Edwards et al. [31] attempted to solve the issue of needing many input variables in order to make energy models viable in residential buildings. Their data was collected from sensors attached to houses, with 140 measurements taken every 15 minutes. The sensors collected data on temperature and time, as well as previous energy consumption readings.

Ambrose et al. [32] explored all aspects of their data for energy efficiency, and how useful each one was for modelling. For example, energy billing data (which will be used for the model of this thesis) can make it hard to determine daily energy patterns if it is calculated on a quarterly basis. Household characteristics like location, BER rating, size etc., as well as income and number of residents can be used as strong input variables for a model. Attributes like indoor temperature, smart sensors readings and what kind of lighting appliances are used for example are impossible to determine without survey on site, and therefore are beyond the scope of this analysis.

Mason et al. [33] took the approach of using monthly data readings to predict the next month's energy usage, which is more applicable to the data that will be explored in this thesis. The analysis of Satre-Meloy [34] found that household electricity use is best described through socio-demographic and physical dwelling variables, like size of the home, and ownership of an electric vehicle, for example. Finally, it was found that the conversion of research from commercial to residential buildings is not straight forward, as the usage patterns can vary quite dramatically.

RQ_2 : **How should machine learning be applied to best predict customer energy efficiency?**

The data needs to be processed before any modelling can be done effectively. Xiao et al. [35] examined the effect of splitting data according to days. For example, Monday energy usage data was used only to predict the following Monday's usage, taking into account holidays where electricity usage would be sporadic. Interestingly, they also used forecasting to predict historical data, in order to prove its validity.

Zekić-Sušac et al. [36] made use of variable reduction using χ^2 tests of independence for the factor variables and correlations for the numeric variables. Based on the initial 47 variables chosen in the sample, this process determined 10 relevant input variables. The factor variables were mapped to binary categories prior to modelling. Zhu et al. [37] used the method of PLS to find the most relevant input variables, while in a lot of cases other simpler methods like normalisation and outlier removal were utilised [38; 39]. Zhang et al. [40] made use of PCA alongside a neural network in order to encapsulate the most important inputs in the most concise and efficient form possible.

Machine Learning in Banking

Machine learning is only useful if it has a clear purpose. In the context of banking and sustainability, this purpose needs to be defined. We can classify customers who are likely to go into default on a loan and help them before this arises using predictions, map geospatial patterns and hence predict customer needs based on their stage in the customer journey. Using this as a basis, the purpose of this analysis is to help customers in their life journey, while also making it as efficient and sustainable as possible. Have they a need for a new car? We can offer them a green lending product for an electric car. Are they buying their first home? We can offer them a green mortgage that give them a good rate of interest, provided the house in energy efficient. These are just samples of what machine learning has done in the past for banking customers and where they can be applied in the future [41; 42; 43].

Liang et al. [44] explored the effect of sustainable action on banking cost efficiency and found that banks who embrace sustainable actions like green products and reducing their carbon footprint outperform those that do not embrace these actions.

Following from this, Taneja et al. [45] explored customer sentiment towards banks taking up sustainable products. By being transparent and showing their intent for action, banks gain approval from their customers. By creating products like the ones mentioned previously, and marketing them in a successful manner using machine learning models to determine which customers would be most open to such a product, its a perfect integration with existing structures such as prediction of credit scores, default prediction etc. This all will lead to a clear definition of energy efficiency in the context of this thesis.

Feature Engineering and Extraction

Not all data is relevant for modelling the dependent variable and must be narrowed down. Chou et al. [46] proposed a framework in which only 4 input variables (outdoor temperature in $°C$, day of the week, hour of the day and the times at which these values were recorded) were taken to create a number of prediction models, using smart sensors in residential housing. Their results indicated that a hybrid model significantly outperforms single models (by 64%), such as support vector machines and linear regression, when using limited data.

Zekić-Sušac et al. [36] in contrast proposed a model structure with 47 input variables (34 continuous and 13 categorical). They devised a preprocessing technique and gathered a sample of ≈ 1500 buildings to train a model, created using a random forest integrated with the Boruta algorithm for variable reduction. They concluded that even with the best model found after cross validation, the accuracy was too low using this minute a data source and required a big data framework in order to get promising results. They determined that the most important input variables for their successful model were: occupational, heating and time data (tying into RQ_1).

Zhao et al. [47] used the correlation coefficient between each variable in their data with the dependent variable and chose which were relevant. They also performed RGS, which was developed by Navot et al. [48] for use in study the study of brain neural activities.

Model Selection

Model selection is crucial for accuracy, and is unique to the dataset at hand. Edwards et al. [31] found, using the environmental sensors, a LSSVM performed best on residential properties. McLoughlain et al. [49] instead used time series forecasting, specifically Fourier transforms and Gaussian processes, and found both performed very well with the variable nature of electricity usage. Mason et al. [33] explored the usage of ANNs, as they assume less about a dataset than other models do. This combined with the monthly data format proved very effective in comparison to state-of-the-art models. Similarly, CNNs, ELMs, FLNNs and GAs were all applied in use cases for commercial properties, but could be altered to suit new data [50; 51; 52; 40].

RQ_3 : **How can users effectively interact with and gain insight about the customer base from this customer model?**

Visualisation

Sacha et al. [53] explored integrating data visualisation techniques with machine learning models. They designed a framework in which human interactions were interwoven with machine learning models, so that results could be made more coherent, as well as allowing the changing of parameters to explore what happens on a dynamic basis.

Hadley Wickham's book, Mastering Shiny [54] provides a comprehensive insight into how to produce interactive dashboards and also how to deploy them for external usage, which would be very relevant to any model plots produced in this thesis. The ability for someone to interact with data that is directly linked to them is beneficial in the adoption and comprehension of findings from the analysis [55].

Ruff et al. [56] explored the use of a shiny dashboard (R language) to visualise the results of a CNN, to make the model accessible to colleagues with less experience in the field of modelling.

Data Processing: What methods can be used to convert spend data to energy data?

Shibano et al. [57] devised a model to convert household income into energy usage data, by linking income to the number of electrical appliances owned and hence, the demand of these appliances, which they deduced to have followed a gamma distribution for simplification. Greer [58] provides case studies in energy cost modelling that mimic real world modelling scenarios. Geo-spacial data could prove very useful when mixed with cost data in determining energy usage by regional means to give a larger view than a single household. Both can be used in conjunction to provide dual perspectives for consumers [59].

Note

The review work by Narciso et al. [30] is the perfect foundation for any paper on this topic and will serve as a guide for a lot of the work to come in this thesis.

The thesis will make a unique contribution to financial sustainability in that (to the best of my knowledge) it will be the first of its kind to investigate green products to Bank X customers through the use of artificial intelligence. The work of Shibano et al. [57] also should be noted as a key area of conversion for feature engineering.

3.3 Review Conclusions

Overall, this review concludes that there is a plethora of evidence to support my hypothesis that energy usage for customers of Bank X can be modelled using spend data as opposed to raw energy data. At the outset of this topic, 4 questions were set to be answered:

1. What are the strengths, weaknesses, threats and opportunities of AI in the field of energy?
2. What are the challenges with energy efficiency modelling?
3. What value or benefits can be achieved by AI in energy efficiency modelling?
4. Can we model customer energy consumption using cost data?

This research is important as AI can be used to model energy usage, whether it be for residential or commercial properties. The caveats however are that usage patterns are a big factor and therefore the same model cannot be used for both types of properties, unless something is done to mitigate this. AI presents us with the opportunity to allow customers of Bank X to see how much they are currently spending and how much they are expected to spend, and compare this to averages for households similar to their own, or on a regional basis. In addition, this has the potential to showcase the minimum energy a household requires/needs. Likewise, business customers who are aiming to be carbon-neutral can work with the bank to forecast where they need to make changes to reach this goal by a certain date.

The challenges surrounding the collection and processing of data for modelling are numerous, but , using techniques outlined in this review, can be minimised. For example, the data collected will be billing data, so this will be in the form of monthly, quarterly, bi-annually or annually and therefore sensor measurements, while providing a good level of background on model input variables and techniques, will be of no benefit to my analysis.

AI can automate tasks that would take an incalculable amount of time for Bank X staff to complete. The model and dashboard could be maintained and updated for new data incoming with little effort. The dashboard would promote helping save the environment as well as saving the customer money in an approachable and simple way.

In conclusion, customer spend data can be used to create a model (using AI) and dashboard (using Shiny) that can predict energy spend, show this spend to the relevant bodies, what customer spend is versus other households similar to their own, and finally tips on how to reduce this energy usage to reduce cost but also benefit the environment, or see what product would be best suited to their needs within the suite of green offerings, all in a simple and efficient manner. There are no studies in the literature researched that do this.

Chapter 4

Data Processing & Analysis

Data Processing

- Describe the dataset I have obtained from Bank X and any other sources.
- What steps did I take in order to make the data clean for production?
- Feature Engineering and Extraction - Any new features developed from other features?
- Bank X data stream - highlight the access to high quality levels of data

Analysis

- Model choices and reasoning
- Models tested
- How the models were created and thus optimised

Chapter 5

Experimental Setup

- Go through the process of how the data was gathered (leaving out the names of key systems and proprietary info etc.).
- Go through how the data was then imported and refined.
- Modelling techniques used, fine-tuning of parameters using cross validation etc.
- Visualisation of things like ROC curve for regression, elbow plot for k means learning etc.
- Production of dashboard of results.
- GitHub repo, but this may not be possible (Can always be made private for Bank X).

Chapter 6

Results

Results first, using figures and tables, with little commentary and no interpretation. Then analysis and interpretation. Statistical tests are typically required to support claims such as superiority of one algorithm over another.

- Sample demo of the dashboard
- Tables of performance metrics etc for each model tested, for before and after cross validation, as well as with different validation set sizing.
- Analysis of these results on a factual level.
- Are the results statistically significant? - χ^2 testing etc.

Chapter 7

Discussion

This chapter is all about my interpretation of the results. While the previous chapter delves into what the results actually **are**, here I can express my own views on what these results actually **mean** in the context of the work I am doing.

- What do the results mean in this context? Not to be confused with a measure of how good or bad they are, but more what these results mean in the real world; do they benefit a customer?
- How the results actually make sense in a real world business case.
- Were they what we expected them to be?
- Where am I surprised at what has been shown? Where am I not?
- Why did some models fail where others succeeded?

Chapter 8

Conclusion

- What have I achieved?
- Where have things gone wrong? Where have they gone better than expected?
- Have I answered the research questions I set at the outset?
- What is different to other study's done on this topic?
- How will this now impact things going forward for Bank X customers?

References

[1] O. Brannock, "MLA3," 08 2021. [Online]. Available: https://github.com/OisinB-2814/MLA3 vi, 13

[2] A. Kaya and M. Keyes, "Energy management technology in pulp, paper and allied industries," *IFAC Proceedings Volumes*, vol. 13, no. 4, pp. 609–622, 1980, 4th IFAC Conference on Instrumentation and Automation in the Paper, Rubber, Plastics and Polymerisation Industries, Ghent, Belgium, 3-5 June 1980. 1

[3] H. Kang, M. Lee, T. Hong, and J.-K. Choi, "Determining the optimal occupancy density for reducing the energy consumption of public office buildings: A statistical approach," *Building and Environment*, vol. 127, pp. 173–186, 2018. 1

[4] A. Roslizar, M. A. Alghoul, B. Bakhtyar, N. Asim, and K. Sopian, "Annual energy usage reduction and cost savings of a school: End-use energy analysis," *The Scientific World Journal*, vol. 2014, pp. 310–539, Nov 2014. 1

[5] J. A. Rodger, "A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings," *Expert Systems with Applications*, vol. 41, no. 4, Part 2, pp. 1813–1829, 2014. 1

[6] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M. Assimakopoulos, R. Mitoula, and S. Zerefos, "Using intelligent clustering techniques to classify the energy performance of school buildings," *Energy and Buildings*, vol. 39, no. 1, pp. 45–51, 2007. 1

[7] D. Torregrossa, U. Leopold, F. Hernández-Sancho, and J. Hansen, "Machine learning for energy cost modelling in wastewater treatment plants," *Journal of Environmental Management*, vol. 223, pp. 1061–1067, 2018. 1

[8] T. W. Edgar and D. O. Manz, "Chapter 6 - machine learning," in *Research Methods for Cyber Security*, T. W. Edgar and D. O. Manz, Eds. Syngress, 2017, pp. 153–173. 4

[9] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 631–641, 2019, 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019. 5

[10] S. Lee, Y. Kim, H. Kahng, S.-K. Lee, S. Chung, T. Cheong, K. Shin, J. Park, and S. B. Kim, "Intelligent traffic control for autonomous vehicle systems based on machine learning," *Expert systems with applications*, vol. 144, p. 113074, 2020. 5

[11] I. P. Adegun and H. B. Vadapalli, "Facial micro-expression recognition: A machine learning approach," *Scientific African*, vol. 8, p. e00465, 2020. 5

[12] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying crisp-dm process model," *Procedia Computer Science*, vol. 181,

pp. 526–534, 2021, cENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020. 6

[13] R. Wirth and J. Hipp, "Crisp-dm: towards a standard process modell for data mining," 2000. 6

[14] H. Wickham, "Tidy data," *The American Statistician*, vol. 14, 09 2014. 6

[15] M. Alloghani, D. Al-Jumeily Obe, J. Mustafina, A. Hussain, and A. Aljaaf, *A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science*, 01 2020, pp. 3–21. 8

[16] J. S. Richman, "Chapter thirteen - multivariate neighborhood sample entropy: A method for data reduction and prediction of complex data," in *Computer Methods, Part C*, ser. Methods in Enzymology, M. L. Johnson and L. Brand, Eds. Academic Press, 2011, vol. 487, pp. 397–408. 8

[17] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966. 10, 12

[18] F. Chollet, *Deep Learning with Python*, 1st ed. USA: Manning Publications Co., 2017. 11

[19] W. L. Dunn and J. K. Shultis, "5 - variance reduction techniques," in *Exploring Monte Carlo Methods*, W. L. Dunn and J. K. Shultis, Eds. Amsterdam: Elsevier, 2012, pp. 97–132. 13

[20] W.-Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Re-*

views: Data Mining and Knowledge Discovery, vol. 1, pp. 14 – 23, 01 2011. 14

[21] X. Chen, Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong, and H. W. Ngan, "Electricity price forecasting with extreme learning machine and bootstrapping," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2055–2062, 2012. 14

[22] M. Awad and R. Khanna, *Support Vector Regression*. Berkeley, CA: Apress, 2015, pp. 67–80. 16

[23] J. Wang, Q. Chen, and Y. Chen, "Rbf kernel based support vector machine with universal approximation and its application," in *Advances in Neural Networks – ISNN 2004*, F.-L. Yin, J. Wang, and C. Guo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 512–517. 16

[24] A. Prieto, B. Prieto, E. M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, and I. Rojas, "Neural networks: An overview of early research, current frameworks and new challenges," *Neurocomputing*, vol. 214, pp. 242–268, 2016. 17

[25] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018. 18

[26] E. Guresen and G. Kayakutlu, "Definition of artificial neural networks with comparison to other networks," *Procedia Computer Science*, vol. 3, pp. 426–433, 2011, world Conference on Information Technology. 18

[27] S. Ruder, "An overview of gradient descent optimization algorithms," 2017. 18

[28] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021. 19

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997. 21

[30] D. A. Narciso and F. Martins, "Application of machine learning tools for energy efficiency in industry: A review," *Energy Reports*, vol. 6, pp. 1181–1199, 2020. 25, 30

[31] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy and Buildings*, vol. 49, pp. 591–603, 2012. 25, 28

[32] M. Ambrose and M. James, "Dealing with energy efficiency data," *Energy Procedia*, vol. 121, pp. 158–165, 2017, improving Residential Energy Efficiency International Conference, IREE 2017. 25

[33] K. Mason, J. Duggan, and E. Howley, "Forecasting energy demand, wind generation and carbon dioxide emissions in ireland using evolutionary neural networks," *Energy*, vol. 155, pp. 705–720, 2018. 26, 29

[34] A. Satre-Meloy, "Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models," *Energy*, vol. 174, pp. 148–168, 2019. 26

[35] L. Xiao, J. Wang, X. Yang, and L. Xiao, "A hybrid model based on data preprocessing for electrical power forecasting," *International Journal of Electrical Power Energy Systems*, vol. 64, pp. 311–327, 2015. 26

[36] M. Zekić-Sušac, A. Has, and M. Knežević, "Predicting energy cost of public buildings by artificial neural networks, cart, and random forest," *Neurocomputing*, vol. 439, pp. 223–233, 2021. 26, 28

[37] L. Zhu and J. Chen, "Energy efficiency evaluation and prediction of large-scale chemical plants using partial least squares analysis integrated with gaussian process models," *Energy Conversion and Management*, vol. 195, pp. 690–700, 2019. 26

[38] Y. Han, C. Fan, M. Xu, Z. Geng, and Y. Zhong, "Production capacity analysis and energy saving of complex chemical processes using lstm based on attention mechanism," *Applied Thermal Engineering*, vol. 160, p. 114072, 2019. 26

[39] B. Beisheim, K. Rahimi-Adli, S. Krämer, and S. Engell, "Energy performance analysis of continuous processes using surrogate models," *Energy*, vol. 183, pp. 776–787, 2019. 26

[40] X.-H. Zhang, Q.-X. Zhu, Y.-L. He, and Y. Xu, "Energy modeling using an effective latent variable based functional link learning machine," *Energy*, vol. 162, pp. 883–891, 2018. 26, 29

[41] J.-P. Li, N. Mirza, B. Rahat, and D. Xiong, "Machine learning and credit ratings prediction in the age of fourth industrial revolution," *Technological Forecasting and Social Change*, vol. 161, p. 120309, 2020. 27

[42] Y. S. Kim and S. Y. Sohn, "Managing loan customers using misclassification patterns of credit scoring model," *Expert Systems with Applications*, vol. 26, no. 4, pp. 567–573, 2004. 27

[43] A. Petukhov, O. Zaikin, and K. Bochenina, "Analysis of the geospatial activity profiles of bank customers," *Procedia Computer Science*, vol. 156, pp.

245–254, 2019, 8th International Young Scientists Conference on Computational Science, YSC2019, 24-28 June 2019, Heraklion, Greece. 27

[44] L.-W. Liang, H.-Y. Chang, and H.-L. Shao, “Does sustainability make banks more cost efficient?” *Global Finance Journal*, vol. 38, pp. 13–23, 2018, special Issue on Corporate Social Responsibility and Ethics in Financial Markets. 27

[45] S. Taneja and L. Ali, “Determinants of customers’ intentions towards environmentally sustainable banking: Testing the structural model,” *Journal of Retailing and Consumer Services*, vol. 59, p. 102418, 2021. 27

[46] J.-S. Chou and D.-S. Tran, “Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders,” *Energy*, vol. 165, pp. 709–726, 2018. 28

[47] H. Zhao and F. Magoulès, “Feature selection for predicting building energy consumption based on statistical learning method,” *Journal of Algorithms and Computational Technology*, vol. 6, pp. 59 – 77, 2012. 28

[48] A. Navot, L. Shpigelman, N. Tishby, and E. Vaadia, “Nearest neighbor based feature selection for regression and its application to neural activity,” vol. 18, 01 2005. 28

[49] F. McLoughlin, A. Duffy, and M. Conlon, “Evaluation of time series techniques to characterise domestic electricity demand,” *Energy*, vol. 50, pp. 120–130, 2013. 28

[50] Z. Geng, Y. Zhang, C. Li, Y. Han, Y. Cui, and B. Yu, “Energy optimization and prediction modeling of petrochemical industries: An improved convolutional neural network based on cross-feature,” *Energy*, vol. 194, p. 116851, 2020. 29

[51] Y.-L. He, P.-J. Wang, M.-Q. Zhang, Q.-X. Zhu, and Y. Xu, “A novel and effective nonlinear interpolation virtual sample generation method for enhancing energy prediction and analysis on small data problem: A case study of ethylene industry,” *Energy*, vol. 147, pp. 418–427, 2018. 29

[52] M. Kovačič and B. Šarler, “Genetic programming prediction of the natural gas consumption in a steel plant,” *Energy*, vol. 66, pp. 273–284, 2014. 29

[53] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim, “What you see is what you can change: Human-centered machine learning by interactive visualization,” *Neurocomputing*, vol. 268, pp. 164–175, 2017. 29

[54] H. Wickham, *Mastering Shiny.* O’Reilly, 2021. 29

[55] Y.-S. Kim, K. Reinecke, and J. Hullman, “Explaining the gap: Visualizing one’s predictions improves recall and comprehension of data,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1375–1386. 29

[56] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan, “Workflow and convolutional neural network for automated identification of animal sounds,” *Ecological Indicators*, vol. 124, p. 107419, 2021. 29

[57] K. Shibano and G. Mogi, “Electricity consumption forecast model using household income: Case study in tanzania,” *Energies*, vol. 13, no. 10, 2020. 30

[58] M. Greer, *Electricity Cost Modelling Calculations - 2nd Edition.* Academic Press, 2021. 30

[59] M. V. Rocco, E. Fumagalli, C. Vigone, A. Miserocchi, and E. Colombo, "Enhancing energy models with geo-spatial data for the analysis of future electrification pathways: The case of tanzania," *Energy Strategy Reviews*, vol. 34, p. 100614, 2021. 30