Bachelor thesis
Computing Science

Radboud University

Data-Driven Estimation of Retail Building Energy Consumption

Author:
Yannick Hogewind
s4413172

First supervisor/assessor:
prof. dr. ir. Djoerd Hiemstra
djoerd.hiemstra@ru.nl

Second supervisor:
Daniël Sies
daniel.sies@vanbeek.com

Second assessor:
dr. ir. Faegheh Hasibi
F.Hasibi@cs.ru.nl

June 28, 2020

Abstract

Commercial buildings represent a substantial part of worldwide energy consumption and greenhouse gas emissions. Modelling this consumption per building allows for the identification of opportunities to increase efficiency, resulting in reduced emissions and costs to businesses. We present a gradient boosted decision tree model that estimates the annual electricity consumption per square meter of floor area for commercial buildings, trained on a dataset containing building features and consumption data for 566 stores operated by a British retailer. The consumption estimates are accurate to within 11.6% mean absolute percentage error, performing similarly to a previously used benchmark on this dataset that is based on physical simulation. These estimates will be used in practice to determine the relative performance of buildings in this data set and identify buildings that are likely to benefit from energy saving renovations.

Contents

Chapter 1

Introduction

In the United States, commercial buildings account for 18% of all energy consumption [1]. With growing populations and increasing wealth in developing countries, building energy consumption is only expected to increase in the coming decades. In the context of climate change, this represents a substantial component of global greenhouse gas emissions [2]. Reducing this consumption by improving building energy efficiency is an important method to lower emissions and combat the effects of global warming, while also reducing costs for businesses.

Van Beek Ingenieurs is a company that aims to accomplish this reduction by providing data monitoring solutions for the energy consumption of commercial buildings, consolidating various sources of electricity, gas and water consumption data. Like other companies in the data monitoring sector, they ensure high data quality and help their clients draw actionable conclusions from this data. Specifically, they identify opportunities to increase building energy efficiency, such as changing the schedules of heating and cooling equipment or installing solar panels. Part of this process is determining whether a building's efficiency is better or worse than expected. Currently, Van Beek uses two methods for benchmarking. Existing buildings are evaluated by comparing their energy consumption with their consumption in previous years; however, this is mostly useful for counteracting consumption increases and does not allow identification of inefficiencies that were already present before data collection started. It also is not possible for newly opened stores, where no past consumption data is available.

An alternative solution to this is to model building energy consumption and compare the expected energy consumption with a building's real consumption to calculate a benchmark score. A possible technique is to use physical modelling methods, but these require extensive architectural information [3] which is not always available and can be expensive to perform.

This is the second benchmarking method used by Van Beek, where physical modelling is used to calculate an expected energy consumption for a portion of newly opened stores.

In this paper, we use a data-driven approach to estimate electrical energy consumption for stores of a British retailer over the previous year. These estimates are calculated using a model which uses only building features as inputs, such as year of construction, sales floor area, refrigerator count and opening hours. The main question answered in this paper is: *How can the energy consumption of commercial real estate buildings be estimated, based on building features using a data-driven method?* To aid in answering this question, we also investigate the following questions: *What applicable data-driven methods exist for prediction of commercial real estate energy consumption?* and *How does the implemented data analysis method perform compared to Van Beek's previous methods?*. The main contribution of this paper is that we show good performance of decision tree ensembles for prediction of annual energy consumption on a real, not simulated, dataset. While a direct comparison is difficult, our method has a lower error than Van Beek's previous benchmark, which involves extensive physical modelling.

The structure of this thesis is as follows. Section 2 contains preliminary knowledge required to understand this thesis, explaining the field of building energy efficiency and the data analysis techniques that are applied in this paper. In section 3 we describe the building data set and the methodology used to develop and evaluate predictive models based on this data set. The resulting models and their performance are described in section 4. In section 5, we give an overview of the state of the art in similar research. In section 6, we present our conclusions.

Chapter 2

Theoretical framework

2.1 Building energy consumption

Buildings consume energy in various forms. The most significant of these are electricity and natural gas, which are used for different purposes: electricity is mostly used for lighting, cooling, powering machinery, ventilation and some space heating, while natural gas is used for space heating. The total energy consumption of a building is the sum of energy consumed in a building in all of these forms. Energy conversion and delivery can be very inefficient; delivering 1 kWh of electrical energy to a building can require over 3 kWh of energy in another form, such as fossil fuels [4]. As such, a distinction is made between *site* energy consumption, i.e. the amount of energy used on-site in a building, and *source* energy consumption, the amount of energy required to generate the on-site consumed energy. For this paper the source consumption is out of scope and we will only discuss the site energy consumption.

2.1.1 Comparing energy efficiency

The energy consumption of buildings is usually specified as energy use intensity (EUI), the energy consumption per unit of floor area in kWh/m^2. This normalizes for the size of a building and provides an indication of the energy efficiency of a building, allowing it to be compared to other buildings. Rating building energy performance by comparing its consumption against other buildings is commonly referred to as benchmarking. Benchmarking is useful as it helps to identify underperforming buildings, which can then be inspected more closely to determine the cause of the low efficiency and remedy it [5].

2.1.2 Energy efficiency in retail buildings

Retail buildings represent 26% of non-domestic electricity usage in buildings in the UK. This consumption can be categorized by various end uses that vary depending on the type of store. For large non-food stores, the median electrical EUI is 146 kWh/m^2, while this is 387 kWh/m^2 for large food stores. This difference is largely explained by the energy required to refrigerate food, which amounts to 220 kWh/m^2 on average in food stores, equal to 57% of the total electricity consumption. In the retail sector as a whole, cooled storage amounts to 32% of all electricity consumption, followed by lighting (26%), space heating (16%) and cooling and ventilation (8%). The remaining 18% is consumed for other end uses, such as IT equipment and machinery [6].

Accordingly, efficiency is mostly determined by the efficiency of these largest end-uses. Heating and space cooling consumption can be reduced by using better insulation material or tuning ventilation schedules; food refrigeration consumption can be reduced by ensuring customers do not open display refrigerators for longer than necessary.

2.2 Regression analysis

Regression is a type of data analysis that models the relationship between some independent variables - also called features or predictors - and a dependent or "outcome" variable [7]. In the context of building efficiency, regression analysis can be used to predict an expected energy consumption for a building. The independent variables could be properties of a building, such as the number of floors it has and the number of refrigerators inside it. A logical choice for the dependent variable could be e.g. the electrical EUI. A regression model that is created based on these variables would then make a prediction of a building's EUI, given its floor count and refrigerator count.

2.2.1 Linear regression

In linear regression, the relationship between the independent variables and the dependent variable is modeled as an equation of the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon$, where y is the predicted value of the dependent variable, β_0 is the y-intercept, the β_i are parameters of the model, the x_i are explanatory variables and ϵ is the *error*, the fluctuation in y that is unexplained by our model [8]. Note that the x_i variables may be the observed independent variables, but they can also be transformed versions of those variables - the model is linear in the *parameters*, not in the independent variables. To continue our previous example where we use floor count and refrigerator count to predict electrical EUI, we might produce the following linear model: $EUI_{electric} = \beta_0 + \beta_1 floors + \beta_2 floors^2 + \beta_3 refrigerators + \epsilon$. The goal is then to find values for β such that ϵ, the error, is minimized for all observations. Different approaches exist to accomplish this.

The most common form of linear regression is ordinary least squares (OLS), in which the model aims to minimize the sum of the squared errors.

While fast to create and easy to understand, linear regression can only be accurate if the relationship it models can be expressed as an equation that is linear in some variables, which is often insufficient. Furthermore, OLS linear regression is sensitive to outliers [8].

2.2.2 Decision trees

Decision tree regression [9] models the relationship between independent and dependent variables using a tree structure not unlike a flowchart, in which each internal node poses a question about an independent variable and has an edge to another node for each answer to that question. The leaf nodes each contain a single value. This regression tree (RT) model can be used to predict the dependent variable by starting at the root node and following the branches of the tree until a leaf node is reached; the predicted value is then equal to the value contained in the leaf node. A graphical representation of a mock decision tree based on our running example is shown in 2.1.

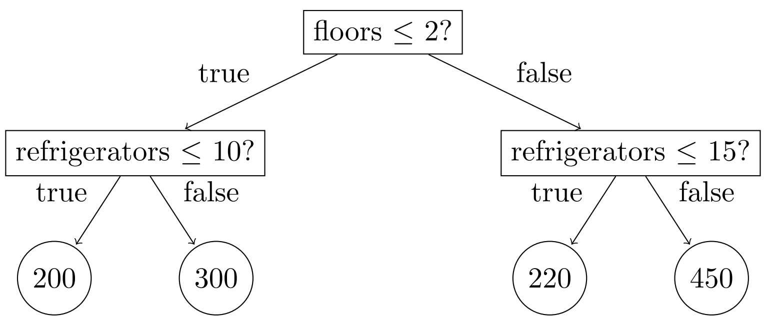

Figure 2.1: A flowchart representation of a mock decision tree that predicts electrical EUI given floor count and refrigerator count.

These can model non-linear relationships and are easy to interpret for humans, as they can be shown as flowcharts and one can see for each prediction exactly which variables caused this decision.

2.2.3 Decision tree ensembles

Ensemble methods aggregate multiple regression models to create one model that is more powerful than its individual components. An extensive overview of ensemble methods, including a theoretical justification for their improved performance, is given in [10]. Three well-known ensemble meta-algorithms that can be applied to decision trees are bagging [11], boosting[12] and random forests [13].

All three operate by using a training set S of size n to learn k decision trees, using a different training set S' also of size n for each individual tree, which is derived from S.

In bagging, S' is created for each tree by uniformly drawing random elements with replacement from S.

Boosting maintains a set of weights over S, and adjusts these weights after learning each tree such that examples with large error receive a larger weight and examples with small error receive a smaller weight; S' is then constructed by drawing elements from S with probability proportional to these weights [14].

Random forests operate like bagging, with the addition that each split decision in a tree may only use a random subset of the independent variables.

2.2.4 Support vector machines

Support vector machines (SVM) for regression aim to find a linear function f that, given independent variables, approximates the dependent variable with an error of at most ϵ. This ϵ-insensitive regression [15] thus ignores observations for which the value predicted by f is within ϵ of the actual value. An f is then found such that the sum of all errors that are greater than ϵ is minimized. The resulting regression generalizes well, as the ϵ margin prevents the algorithm from overfitting to patterns in small, random noise. While SVM regression learns a linear function, it is typically used to model non-linear relations through use of a kernel: a helper function that projects the independent variables to another space, such that the regression function to be found is linear in that space. For an excellent introduction to support vector regression, we refer the reader to the 2004 tutorial by Smola and [16].

2.3 Summary

In summary, linear regression, decision tree ensembles and support vector machines are popular methods for regression analysis. While linear regression models are simple to develop and interpret, they are unable to accurately model non-linear relations between variables. In addition, they are sensitive to outliers. Decision tree ensembles and support vector machines are more difficult to interpret, but robust to outliers and well suited for modelling non-linear relations. In the rest of this thesis we will investigate how these regression analysis methods can be applied to model building electricity consumption.

Chapter 3

Data and Methods

Full source code for the described methods is available at
github.com/yhogewind/Code-BachelorThesis.

3.1 Data description and pre-processing

The available dataset consists of building features and energy consumption in 2019 for 771 buildings owned by a British retailer. The data of interest was extracted from a proprietary energy management platform to a local database using Python 3.7. To focus the analysis on active stores we excluded office buildings, warehouses, and stores that were not open for the entirety of 2019, yielding a total of 566 buildings used for model development and testing. This set of stores is diverse: their size ranges from convenience stores to large clothing shops with attached cafes, and they can sell clothing and home products, food, or both.

For each building a list of assets was available, detailing individual items on site such as bakery ovens, elevators and refrigerators; these were aggregated to a count of items per category per store. The target variable, EUI, was calculated by dividing the annual electric energy consumption by the total internal floor area of the store. In addition, the ratio between floor area dedicated to food and the total floor area was calculated, as previous research has shown this to be strongly correlated with EUI [17]. Each store also had a type associated with it, indicating whether it is mostly food-based or not and the physical placement: in a retail park, shopping mall or a city center. A summary of the dataset after these transformations is provided in table 3.1.

3.2 Model development

Before model development the data was randomly split into a section of 20% (n=114), to later be used to test model accuracy and a section of 80%

	mean	std	min	max
Dependent variable				
EUI (kWh/m^2)	474.19	237.31	94.92	1350.82
Independent variables				
Food floor area (m^2)	933.82	508.12	0.0	3692.99
Total floor area (m^2)	2698.10	2826.07	163.97	15255.98
Food floor area as fraction of total area	0.60	0.36	0.0	1.0
Presence of clothing sales (binary)	0.095	0.29	0.0	1.0
Presence of toilets (binary)	0.24	0.43	0.0	1.0
Presence of gas supply point (binary)	0.84	0.37	0.0	1.1
Average opening hours per day	12.88	1.87	7.41	24.0
Count of “Catering” assets	22.55	17.06	0.0	107.0
Count of “Air Conditioning” assets	74.53	62.26	0.0	499.0
Count of “Lift” assets	4.91	6.10	0.0	57.0
Count of “Refrigeration” assets	86.16	41.50	0.0	310.0
Store opening year	2004.80	6.86	2000.0	2018.0
Length of refrigeration bays (feet)	320.33	128.58	0.0	841.0

Table 3.1: Summary of the dataset after transformations, excluding encoding of store type.

Parameter	Search space
n_trees	100, 200, 400, 800
max_depth	2, 4, no limit
min_samples_leaf	1, 2, 4
min_samples_split	2, 3, 4

(a) Random forest

Parameter	Search space
n_trees	100, 200, 400
max_depth	2, 4, no limit
min_samples_split	2, 3, 4
learning rate	0.001, 0.01, 0.1

(b) Gradient-boosted trees

Parameter	Search space
kernel	polynomial, rbf
degree	4, 5, 6, N/A
coef0	1, 5, 10, N/A
epsilon	0.001, 0.01, 0.1, 1, 10
regularization C	0.1, 0.2, 0.5, 1.0, 10.0, 100, 1000
tolerance	0.01, 0.1, 0.2

(c) Support vector machine

Table 3.2: Search space for the parameters of each algorithm. Optimal parameters are underlined.

(n=452), used for training and validation. We used the Python library scikit-learn [18] for model and parameter selection. We chose three regression models to develop, as they have shown good performance in similar research on a simulated dataset [19]: random forest (RF), gradient boosted decision trees (GBT) and support vector machine (SVM). In addition, we developed an ordinary least squares linear regression model, which is a simple and popular method for EUI estimation [20]. All algorithms except the linear regression require meta parameters to be specified. As RF and GBT are both decision tree ensembles, they share some parameters: maximum tree depth, the number of trees in the ensemble and the minimum samples in a node required to split that node. The SVM parameters were the type of kernel used for data transformation, the regularization parameter C, the stopping tolerance and epsilon; in addition, if the polynomial kernel is used, the degree and the independent term in the polynomial kernel must be chosen. An overview of the search space for these parameters is shown in table 3.2. The optimal parameters were selected using grid search with 10-fold cross-validation on the training-validation set.

We then used the optimal parameters for each algorithm to create a model, trained on the entire training-validation set.

3.3 Evaluation

After developing the four models, we evaluated their performance using the previously held-out testing dataset. Three metrics were used to compare the error between the predicted consumption values and the actual consumption: mean absolute percentage error, coefficient of determination and the percentage of margin around the predicted values that contains 80% of the actual values. These metrics are chosen as they are common in other annual EUI estimation studies [21] and are easy to understand.

3.3.1 Mean absolute percentage error

The mean absolute percentage error (MAPE) is defined for n predictions and actual values as the equation below [21]. It reflects the average prediction error as a percentage of the actual value, which is an easily understood measure of the prediction accuracy.

$$MAPE = \frac{1}{n} \sum_{i}^{n} \left| \frac{y_{predicted,i} - y_{actual,i}}{y_{actual,i}} \right| * 100\%$$

3.3.2 Coefficient of determination (R^2)

The coefficient of determination is a well-known metric in statistics that represents the portion of the variance in a dependent variable that is explained by an independent variable. Choosing the predicted EUI as the dependent variable and the actual EUI as the independent variable, R^2 provides an indication of the strength of our predictions. An R^2 equal to 1 would mean that the model under test performs perfectly, while an R^2 equal to 0 shows that the model performs no better than if it always predicted the mean of the target variable. R^2 is defined for n predictions and actual values as the equation below, in which $\overline{y_{actual}}$ is the mean of the actual y values [21].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{predicted,i} - y_{actual,i})^2}{\sum_{i=1}^{n}(y_{actual,i} - \overline{y_{actual}})^2}$$

3.3.3 80th percentile margin

The 80th percentile margin is used to evaluate energy consumption estimation in [22], to provide an indication of the percentage margin required around predicted values to be reasonably sure the actual value will be within the margin. This is calculated by first determining the absolute error for each prediction, as a percentage of the predicted value: $\left| \frac{y_{predicted} - y_{actual}}{y_{predicted}} \right|$ and choosing the value at the 80th percentile.

Chapter 4

Results

The trained models were evaluated using the data previously set apart for testing. The performance metrics for each model on this testing set are shown in table 4.1. These show that all models make a prediction of EUI with an error of under 16% on average, and a high R^2 indicates that much of the variation in EUI is explained by all models. It should be noted that the models predict EUI, which is by definition the total consumption divided by floor area. As there is a strong correlation between total consumption and floor area, the models attain a higher R^2 if used to predict total consumption instead: our exploratory tests show that the R^2 when predicting total energy consumption is 0.94 for the random forest model; however, the MAPE is identical.

Figure 4.1 shows the predicted EUI, plotted against the actual EUI for each building in the testing set. Immediately apparent in each graph is a cluster located between an actual EUI of 200 and 400 kWh/m^2. This cluster almost exclusively contains stores that have a small portion of floor area dedicated to food sales and thus consume little energy for refrigeration and cooking. The linear model performs the worst of all four models. In the linear model, this cluster between 200 and 400 kWh/m^2 is less tightly grouped than in the other models; in fact, there is more spread across the entire range of EUI values. This is visible in the graph and it is also reflected in the high MAPE (15.8%) and 80% margin (22.4%). The graph also shows one prediction near zero, where the actual consumption is closer to 100, while all other models provide accurate estimations even in this low range.

The second worst performance is shown by the SVR, which is marginally less accurate than the tree ensembles. In the high end of the EUI these estimates are substantially lower than the actual consumption, which can also be seen in the linear regression. While this bias does indeed represent inaccurate estimates in this range, an argument could be made that the stores with such high EUI might be inefficient and therefore should have a lower EUI. If that is the case, this underestimation of high EUI values is

	MAPE	R^2	80% margin
OLS linear model	15.8%	0.90	22.4%
Random forest	11.7%	0.91	16.8%
Gradient boosted trees	11.6%	0.90	17.6%
Support vector machine	13.5%	0.90	19.6%

Table 4.1: Performance metrics on the test data for each developed model.

appropriate for a model that is to be used for benchmarking by comparing estimates to real values.

The tree ensemble models, random forest and gradient boosted trees, yield the most accurate predictions of all four models. Across the entire range of EUI values they provide accurate estimates, which is reflected in the low MAPE. Of the two, the random forest model appears to be slightly better with a lower 80% margin, which indicates a lower spread.

4.1 Comparison with physical benchmark

The physical modelling approach previously used by Van Beek has been applied to 26 stores opened between 2015 and 2019. It is difficult to make a fair comparison between this previous approach and our models, for several reasons. The stores which have been modelled physically are not all contained in our randomly selected test set; as such, we cannot compare predictions for the same buildings. Secondly, the physical simulations were calculated over different periods than our analysis, which exclusively used data from 2019. Most importantly, the physical simulation estimates are predictions for a future period, whereas our methods provide an estimate of the past year's consumption. This gives our methods an advantage as they are not subject to uncertainty due to future weather or changes in consumer behaviour.

Still, we can make some valuable comparisons to assess whether the performance of our models is acceptable for use by Van Beek. The previous physical simulations had a MAPE of 20.1%. The 11.6% MAPE shown by our best model is significantly lower, indicating that its estimation accuracy is sufficiently high to be of value in Van Beek's use cases. In addition to this improved accuracy, our methods are easy to apply to any building in the data set. While the physical simulations require extensive architectural data that is not available for all buildings and requires much preparation per prediction, our methods work using readily available data and can be applied to all buildings with little effort. As such, our models are deemed to be suitable for use by Van Beek and will be used alongside the physical modelling for new stores as well as on all existing stores to assess energy performance.

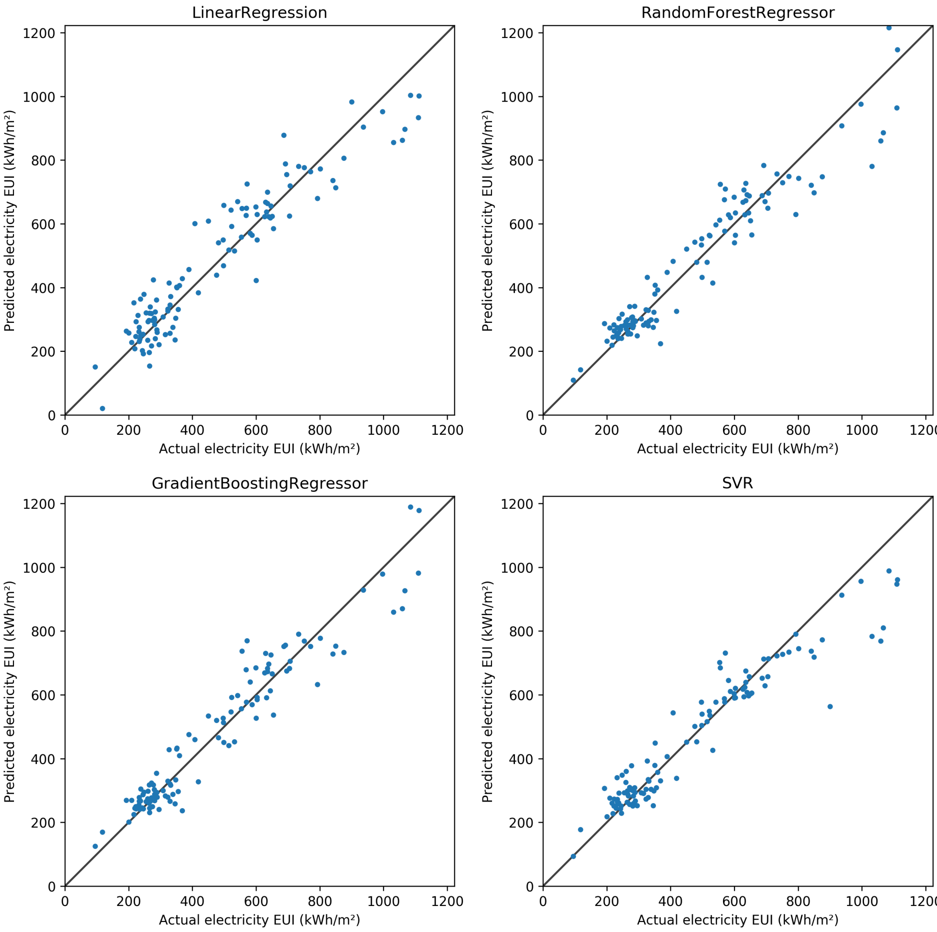

Figure 4.1: Actual electrical EUI vs estimated EUI for the testing set. The black line indicates x=y.

Chapter 5

Related Work

Several recent literature reviews exist for building energy consumption benchmarking [23], [24], [20], [3]. A selection of studies discussed in these reviews is described below.

Sharp [25] was first to estimate EUI based on building features for benchmarking purposes, using OLS linear regression. This approach was recently used to predict real annual electricity consumption in a data set of 188 UK supermarkets, reaching an R^2 of 0.75 [17].

Artificial neural networks (ANN) have been used for EUI prediction with varying rates of success [3]. Zhang finds that ANNs are less effective than simple linear regression models [26], as do Li and Huang [27]. In contrast, ANNs can perform well when forecasting hourly energy consumption on a simulated data set [22], and are used to develop effective benchmarks on real data in [28] and [29].

Support vector machines (SVM) are another commonly used method for energy consumption prediction. An SVM is used by Li [30] to predict the annual EUI of 59 residential buildings, using a dataset of real energy consumption measurements and extensive architectural data. The SVM model is compared with various ANNs and is shown to have the lowest training and testing errors. In [31], SVMs are used effectively for short- to medium-term prediction of real measured energy consumption, given energy consumption in previous time periods.

Tree-based ensemble methods are used for energy consumption estimation and shown to be capable of outperforming state-of-the-art methods in [19]. This study compares tree-based ensemble models: random forest, extremely randomized trees and gradient-boosted regression trees to state-of-the-art methods SVM, SVM+ANN ensemble [32] and genetic programming [33].

These different techniques are compared by applying them on a shared data set. This data set uses physical modelling techniques on hypothetical buildings with random architectural features such as surface area and height to obtain a simulated heating and cooling load for each building. The tree ensembles perform best, with gradient-boosted regression trees improving state-of-the-art predictions of heating load by 14% and cooling load by 65%, while also requiring the least computational power to train.

Quantile regression is introduced as a benchmarking tool in [34], by creating a different quantile regression model for each of 92 percentiles. A building's benchmark score is calculated by applying each regression model to the building's features and choosing the model which has the predicted EUI closest to the building's actual EUI. The benchmark score is then equal to 100 minus the percentile of that model, giving an indication of the relative performance of the building on a scale from 0 to 100. No assessment of accuracy is given aside from a test of consistency of benchmark scores between buildings in two subsequent years. The dataset used in this study contains only building features and real electricity and gas consumption values.

Some publicly available data sets exist that could be used to compare methods for building energy consumption estimation, but we decided not to use these. The Commercial Buildings Energy Consumption Survey data (CBECS) [35] is the basis for the United States government energy benchmarking tool, EnergyStar. This data set was not used in this paper due to its small sample size and low quality [36]. The ASHRAE great energy predictor shootout data set [37] is based on real consumption measurements and is of high quality, but is intended for time-series, short-term forecasting of consumption based on historical consumption and as such is not appropriate for our intended model goals.

5.1 Summary

The literature indicates that many different algorithms can be applied successfully in the space of building energy consumption estimation. Tree ensemble models appear to be the strongest method for annual estimations that do not depend on historical energy consumption, while support vector regression and neural networks perform well on short term and high resolution time-series forecasting. It should be noted that many of these studies are based on either small data sets or data sets based on physical modelling rather than real measured consumption, which may mean the results do not generalize well to other data sets based on real measurements [21][38]. Finally, a few public building energy consumption data sets exist, but these were not used for comparison in our research as they were either of low quality or not intended for the type of estimation performed in our work.

Chapter 6

Conclusions

We have shown that decision tree ensembles can be used to accurately predict electricity consumption for retail buildings. While all of the investigated methods performed adequately, linear regression had the lowest accuracy. Support vector machine, random forest and gradient boosted decision trees all performed well, with the support vector machine performing marginally worse than the two decision tree ensembles, of which random forest was most accurate by a small margin. Our results indicate that the good performance of decision tree ensembles in simulated energy consumption data sets also holds when applied to real energy consumption data. Accordingly, we believe this prediction can be used to inform energy management decisions and achieve a reduction in energy consumption. As the gradient boosted ensemble provided consumption estimations in line with Van Beek's current modelling technique for new stores while being easier to apply, it will be used by Van Beek in practice to evaluate the energy performance of both new and existing buildings.

Bibliography

[1] U.S. EIA. *Monthly Energy Review.* Washington, DC: U.S. Energy Information Administration, Feb. 25, 2020, p. 35.

[2] Luis Pérez-Lombard, José Ortiz, and Christine Pout. "A review on buildings energy consumption information". In: *Energy and Buildings* 40.3 (Jan. 1, 2008), pp. 394–398. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2007.03.007.

[3] E. H. Borgstein, R. Lamberts, and J. L. M. Hensen. "Evaluating energy performance in non-domestic buildings: A review". In: *Energy and Buildings* 128 (Sept. 15, 2016), pp. 734–755. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2016.07.018.

[4] Marek Czachorski and Neil Leslie. "Source energy and emission factors for building energy consumption". In: *American Gas Association, prepared for Natural Gas Codes and Standards Research Consortium* (2009). Publisher: Citeseer.

[5] Luis Pérez-Lombard et al. "A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes". In: *Energy and Buildings* 41.3 (Mar. 1, 2009), pp. 272–278. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2008.10.004.

[6] Department for Business, Energy & Industrial Strategy. *Building Energy Efficiency Survey: Retail sector 2014-15.* Library Catalog: www.gov.uk. Nov. 2016.

[7] Norman Richard Draper and Harry Smith. *Applied regression analysis.* 3rd ed. Wiley series in probability and statistics. New York: Wiley, 1998. 706 pp. ISBN: 978-0-471-17082-2.

[8] G. A. F. Seber and A. J. Lee. *Linear regression analysis.* 2nd ed. Wiley series in probability and statistics. OCLC: ocm51635479. Hoboken, N.J: Wiley-Interscience, 2003. 557 pp. ISBN: 978-0-471-41540-4.

[9] Leo Breiman, ed. *Classification and regression trees.* Repr. OCLC: 247053926. Boca Raton: Chapman & Hall [u.a.], 1998. 358 pp. ISBN: 978-0-412-04841-8.

[10] Ye Ren, Le Zhang, and P.N. Suganthan. “Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]”. In: *IEEE Computational Intelligence Magazine* 11.1 (Feb. 2016). Conference Name: IEEE Computational Intelligence Magazine, pp. 41–53. ISSN: 1556-6048. DOI: 10.1109/MCI.2015.2471235.

[11] Leo Breiman. “Bagging predictors”. In: *Machine Learning* 24.2 (Aug. 1, 1996), pp. 123–140. ISSN: 1573-0565. DOI: 10.1007/BF00058655.

[12] Jerome H. Friedman. “Stochastic gradient boosting”. In: *Computational Statistics & Data Analysis*. Nonlinear Methods and Data Mining 38.4 (Feb. 28, 2002), pp. 367–378. ISSN: 0167-9473. DOI: 10.1016/S0167-9473(01)00065-2.

[13] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001). Publisher: Springer, pp. 5–32.

[14] Thomas G. Dietterich. “An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization”. In: *Machine Learning* 40.2 (Aug. 1, 2000), pp. 139–157. ISSN: 1573-0565. DOI: 10.1023/A:1007607513941.

[15] Vladimir N Vapnik. *The Nature of Statistical Learning Theory*. OCLC: 864225872. New York, NY: Springer New York : Imprint : Springer, 2000. ISBN: 978-1-4757-3264-1 978-1-4419-3160-3.

[16] Alex J. Smola and Bernhard Schölkopf. “A tutorial on support vector regression”. In: *Statistics and Computing* 14.3 (Aug. 2004), pp. 199–222. ISSN: 0960-3174. DOI: 10.1023/B:STCO.0000035301.49549.88.

[17] Maria S. Spyrou et al. “An empirical study of electricity and gas demand drivers in large food retail buildings of a national organisation”. In: *Energy and Buildings* 68 (Jan. 1, 2014), pp. 172–182. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.09.015.

[18] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[19] Sokratis Papadopoulos et al. “Evaluation of tree-based ensemble learning algorithms for building energy performance estimation”. In: *Journal of Building Performance Simulation* 11.3 (May 4, 2018). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/19401493.2017.1354919, pp. 322–332. ISSN: 1940-1493. DOI: 10.1080/19401493.2017.1354919.

[20] William Chung. “Review of building energy-use performance benchmarking methodologies”. In: *Applied Energy* 88.5 (May 1, 2011), pp. 1470–1479. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2010.11.022.

[21] Kadir Amasyali and Nora M. El-Gohary. “A review of data-driven building energy consumption prediction studies”. In: *Renewable and Sustainable Energy Reviews* 81 (Jan. 1, 2018), pp. 1192–1205. ISSN: 1364-0321. DOI: 10.1016/j.rser.2017.04.095.

[22] Alberto Hernandez Neto and Flávio Augusto Sanzovo Fiorelli. "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption". In: *Energy and Buildings* 40.12 (Jan. 1, 2008), pp. 2169–2176. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2008.06.013.

[23] Matheus Soares Geraldi and Enedir Ghisi. "Building-level and stock-level in contrast: A literature review of the energy performance of buildings during the operational stage". In: *Energy and Buildings* 211 (Mar. 15, 2020), p. 109810. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2020.109810.

[24] Yixuan Wei et al. "A review of data-driven approaches for prediction and classification of building energy consumption". In: *Renewable and Sustainable Energy Reviews* 82 (Feb. 1, 2018), pp. 1027–1047. ISSN: 1364-0321. DOI: 10.1016/j.rser.2017.09.108.

[25] Terry Sharp. "Energy Benchmarking In Commercial Office Buildings". In: *Proceedings of the ACEEE*. 1996, p. 9.

[26] Yuna Zhang et al. "Comparisons of inverse modeling approaches for predicting building energy performance". In: *Building and Environment* 86 (Apr. 1, 2015), pp. 177–190. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2014.12.023.

[27] Zhengwei Li and Gongsheng Huang. "Re-evaluation of building cooling load prediction models for use in humid subtropical area". In: *Energy and Buildings* 62 (July 1, 2013), pp. 442–449. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2013.03.035.

[28] A. P. Melo et al. "Development of surrogate models using artificial neural network for building shell energy labelling". In: *Energy Policy* 69 (June 1, 2014), pp. 457–466. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2014.02.001.

[29] Melek Yalcintas. "An energy benchmarking model based on artificial neural network method with a case example for tropical climates". In: *International Journal of Energy Research* 30.14 (2006). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.1212, pp. 1158–1174. ISSN: 1099-114X. DOI: 10.1002/er.1212.

[30] Qiong Li, Peng Ren, and Qinglin Meng. "Prediction model of annual energy consumption of residential buildings". In: *2010 International Conference on Advances in Energy Engineering*. 2010 International Conference on Advances in Energy Engineering. June 2010, pp. 223–226. DOI: 10.1109/ICAEE.2010.5557576.

[31] Elena Mocanu et al. "Deep learning for estimating building energy consumption". In: *Sustainable Energy, Grids and Networks* 6 (June 1, 2016), pp. 91–99. ISSN: 2352-4677. DOI: 10.1016/j.segan.2016.02.005.

[32] Jui-Sheng Chou and Dac-Khuong Bui. "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design". In: *Energy and Buildings* 82 (Oct. 1, 2014), pp. 437–446. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2014.07.036.

[33] Mauro Castelli et al. "Prediction of energy performance of residential buildings: A genetic programming approach". In: *Energy and Buildings* 102 (Sept. 1, 2015), pp. 67–74. ISSN: 0378-7788. DOI: 10.1016/j.enbuild.2015.05.013.

[34] Jonathan Roth and Ram Rajagopal. "Benchmarking building energy efficiency using quantile regression". In: *Energy* 152 (June 1, 2018), pp. 866–876. ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.02.108.

[35] U.S. EIA. *2012 Commercial Buildings Energy Consumption Survey (CBECS) Data*. 2015. URL: https://www.eia.gov/consumption/commercial/data/2012/index.php?view=microdata (visited on 03/09/2020).

[36] John H Scofield. "ENERGY STAR Building Benchmarking Scores: Good Idea, Bad Science". In: ACEEE Summer Study on Energy Efficiency in Buildings. Pacific Grove, California, Aug. 17, 2014, p. 16.

[37] J. F. Kreider and J. S. Haberl. "Predicting hourly building energy use: The great energy predictor shootout – Overview and discussion of results". In: (Dec. 31, 1994). Number: CONF-9406105- Publisher: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States). ISSN: 0001-2505.

[38] Nan Li et al. "Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?" In: *Applied Energy* 159 (Dec. 1, 2015), pp. 196–205. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2015.09.001.