EMISSIONS AND ELECTRICITY USE BY BUILDING

Tracking Energy at WMU

CO2 Emissions on WMU's Campus

We would like to thank those at Western Michigan University (WMU) Facilities Management who helped us in the process of obtaining data for this project. Thank you, Keith Pung (Environmental Health and Safety), for the building-level EPA emissions report data; thank you, George Jarvis (Robert M. Beam Power Plant Director), for the power plant data, various interviews and the power plant tour; thank you, Pete Strazdas (Associate Vice President of Facilities Management), for the interview and for the connection to Joel Ward; and, thank you, Joel Ward (Energy Administration Specialist), for the building-level use data. Also, thank you, Chris Caprara (Energy Systems Specialist), for looking through our results from a greenhouse gas (GHG) accounting perspective.

Arayana Janson, Andrew Messecar, Shelby Mize
"CO2 Emissions Reduction Team"

Abstract

Western Michigan University's (WMU) upcoming budget model transition presents a prime opportunity to consider the introduction of an internal climate price resembling the successful efforts made at institutions such as Yale University[iii]. College deans and other invested stakeholders would be responsible for the decisions leading to the success of such a program. Machine learning algorithms have the potential to simplify the process of analyzing the relevant data in order to assist our university's leaders in the monitoring and evaluation of an internal carbon pricing initiative. Additionally, machine learning's predictive power can help guide the program's development as well as its implications for other university-wide decisions and investments being considered. Multiple linear regression, random forest models, and neural networks were all applied to the analysis of qualitative and quantitative data that is relevant to understanding carbon emissions at WMU. While the most flexible and intricate of our models, neural networks underperformed when compared with other model types, either overfitting the data or indicating an absence of nonlinear relationships in the data. A random forest model demonstrated the highest test performance of all three model types; however, multiple linear regression outperformed the other two model varieties in terms of predictive accuracy when all three were tested identically.

TABLE OF CONTENTS

I. Introduction

I.I Context

The original intention of this project was to investigate CO2 emissions by building for all of WMU's Kalamazoo campus. We have access to the EPA report[i] data on WMU's GHG emissions by building (through Keith Pung); this data reports natural gas use by building (in MCF) to the EPA which can be converted to carbon emissions by building (in Mte). We are not entirely sure either of how gas MCF by building is derived or how the EPA derives CO2 Mte from our MCF numbers, but we are inclined to think that the MCF by building is simply natural gas meter numbers at the building, which are then converted by the EPA to Mte. That said, the gas use or CO2 in MCF by building data set—since we don't have other metering data for all buildings—would not allow us a complete view or accurate breakdown of building-level emissions data related to energy consumption. The natural gas meters at the campus buildings (besides the power plant) are only needed for kitchen cookers or other similar appliances; therefore, gas usage metered at buildings comprises a negligible percentage of CO2 emissions produced at the college.

It was revealed to us that energy consumption by building is comprehensively derived from buildings' electricity and steam meters. In acquiring building-level energy consumption data, we understood that, in the process of requesting data from WMU's Facilities Management, we might encounter discrepancies and gaps—especially since some of the metering data is obtained by student employees.

The general incentive for this project is an impending transformation of the WMU budget model from centralized to decentralized[ii], which means that college deans (or responsibility center leaders) will be responsible for their budget unit's electricity bills (and therefore energy use). The goal of this project is to assist WMU administrators in understanding the availability, integrity, and relevance of data from the power plant and building meters, provided by Facilities Management, that would be used in determining energy bills for budget units. Where Facilities Management has, until now, overseen all campus emissions reductions and energy conservation efforts for all buildings on the WMU campus, those goals will (or could) now be handled by budget unit heads and college deans. We initially hoped that we could build a machine learning model that could be used in assessing the effects of energy use at the building level on CO2 emissions across campus, as well as the outcomes—were CO2 emissions to charged—on budget units' energy bills and on campus emissions. We have succeeded in applying machine learning to our data: a first step towards this goal.

It should be noted that part of the reason we are interested in helping to assess the availability and quality of data is that we know that the access to appropriate information on carbon emissions as well as relevant energy use information has helped other campuses that also contend with a decentralized budget model, not only for billing purposes, but also for reducing energy usage. Due to the availability of more (accurate) energy use and emissions data by building, Yale University's first pilot study on internal carbon pricing showed a 7.4% reduction in emissions—this reduction was achieved without a charge per ton of CO2 emissions (while the overall pilot study was a comparison of different internal carbon pricing mechanisms).[iii]

Aside from the internal carbon price, we have learned (regarding energy billing) that other universities that are on the decentralized budget model have engaged a kind of hybrid energy use tracking system, which uses both MMBtu (for a total usage number by building that reflects monthly steam and electricity use) as well as square footage where MMBtu information is lacking (due to a deficiency in building level meters, or submeter malfunctions for that building)[iv]. About 90-95% of the WMU campus buildings are sub-metered for both steam and electricity.[v] Also, about 90% of WMU campus buildings are connected on the steam system, which is called the 'District Energy System'.[vi] (This group of authors does not yet know if there is 100% correlation between the buildings that are sub-metered and the buildings that are connected to the steam-based District Energy System.)

All of the energy coming out of the power plant (and going to the buildings) can be categorized as either electricity or ("natural-gas-based") steam energy.

I.II Data Gathering

From March 1, 2021 to April 13, 2021, we were in communication with four members of Facilities Management (Pete Strazdas, George Jarvis, Joel Ward, and Dan Brimmer) in the interest of acquiring our data for this project. At first, we hoped that building-level data for energy use by energy source type (steam or electricity) would be available for all buildings. Had we been able to obtain this data for all buildings, we would have then compared the extrapolated, building-level emissions to power plant data on emissions and energy sources. For example, we could have used this data to investigate production and usage efficiencies regarding heat-recovered steam (or, "free steam") versus turbine-generated steam as these relate to numbers for steam that is recovered in the plant's balance condenser (or, unused steam).[vii] We ultimately determined that this analysis would not currently be possible, given our project completion time constraint. However, we will discuss in the conclusion section how our analysis process (and further applications of them) might be relevant in the future, once the full set of building-level data (for the approximate 90% of the campus buildings that are sub-metered[viii]) is acquired.

I.III About the Data

Our final data set of 196 rows is comprised of steam and electricity sub-meter data by building from annual Utility Usage reports for Student Affairs, as compiled by Facilities Management[ix], regarding conference, dining and residence buildings. We compiled data by semester for nine full semesters (from Summer II 2016 to Fall 2019) and on fourteen buildings. Out of nineteen buildings for which data was provided to us, the fourteen selected buildings had the most complete data available on electricity and steam use. (To note, several of these buildings have either been demolished or are schedule to be demolished.[x])

One of our variables requires some explanation, and that is "Building's fraction of Beam Plant emissions". The number associated with the variable here corresponds with our complete variable set chart below.

13. **Building's fraction (called "contribution" in data set) of Beam Plant emissions, Mte:** this variable compares the energy used per building with the total amount of energy generated by the plant to form a percentage of the campus's energy used in a semester by that building. This percentage is then multiplied by the total emissions from the plant to determine the portion or faction of total emissions by semester that can be attributed to the building's energy use.

To arrive at the above variable, we first calculated total energy use by building per semester in MMBTus, in accordance with electricity and steam sub-meter data in kWhs[xi] and klbs[xii], respectively. Also converted to MMBTus was the total energy production by the Beam Power Plant from meters at the outgoing electricity feeders (i.e., power lines) in kWh[xiii] and steam lines (i.e., pipes) in klbs.[xiv] By dividing the former quantity (building energy use) by the latter (power plant energy production), we arrived at the fraction of Beam's total energy used per building per semester. This calculated percentage was multiplied by the Beam plant's total CO2 emissions, giving us a fraction of Beam plant emissions by building.[xv - xvi]

The following table (Table 1) portrays variable names, data types, variable types, and values (for nominal or categorical variables) or ranges (for numerical variables) in our final data set.

VARIABLE NAME	VARIABLE DATA TYPE	VARIABLE TYPE	VALUES (Nominal) or RANGE (Numerical)
1. Building	String	Nominal	"Bernhard", "Burnham Halls", "Draper/Siedschlag", "Valley 1", "Valley 2", "Valley 3", "Valley Dining", "Davis Hall", "Davis Dining", "French Hall", "Henry Hall", "Spindler Hall", "Western Height 175"
2. Electricity use (by building by semester), kWh[xvii]	Numeric	Numerical	Range: 542 kWh – 1,183,537 kWh
3. Steam use (by building by semester), Klbs[xviii]	Numeric	Numerical	Range: 4 Klbs – 59,993 Klbs
4. Semester	String	Nominal	"Summer I", "Fall", "Spring", "Summer II"
5. Year	Numeric	Nominal	2016, 2017, 2018, 2019
6. Building square footage[xix]	Numeric	Numerical	Range: 21,036 SF – 239,464 SF
7. Number of floors in building[xx]	Numeric	Nominal	"2 Floors", "3 Floors", "5 Floors", "6 Floors"
8. Primary use of building[xxi]	String	Nominal	"Conference", "Housing", "Dining"
9. Year built[xxii]	Numeric	Nominal	1940, 1948, 1950, 1954, 1955, 1957, 1960, 1963, 1964, 1965, 2016, 2017
10. Average temperature by semester, degrees F[xxiii]	Numeric	Numerical	Range: 32.5 degrees F – 74.0 degrees F
11. WMU enrollment by semester, all campuses[xxiv]	Numeric	Numerical	Range: 5,317 – 23,252 students
12. Cost of natural gas, dollars per thousand cubic ft[xxv]	Numeric	Numerical	Range: 6.628 $/Th Ft^3 – 9.555 $/Th Ft^3
13. Building's fraction of Beam Plant emissions, Mte [xxvi, xxvii, xxviii, xxix, xxx, xxxi, xxxii]	Numeric	Numerical	Range: .0068 – 3.5127

Table 1: Raw data set variables.

II. Methods and Results

II.I Feature Selection

With our data compiled, we began exploring the relationships between the different parameters of our data set. One of the first useful steps to take was to check our variables for the occurrence of multicollinearity. We did this by observing the Variance Inflation Factor (VIF) for each variable. We used the VIF to identify or confirm and then abandon highly collinear variables for the construction of our models—since VIF detects standard errors and therefore variances that are inflated due to multicollinearity.

In additional processes, we further reduced our number of active variables during prep stages for our various model types. For example, for our first model type, which was multiple linear regression, the relationships between the remaining variables were explored using a correlation matrix (Figure 1 below) in conjunction with a series of box plots and scatter plots.

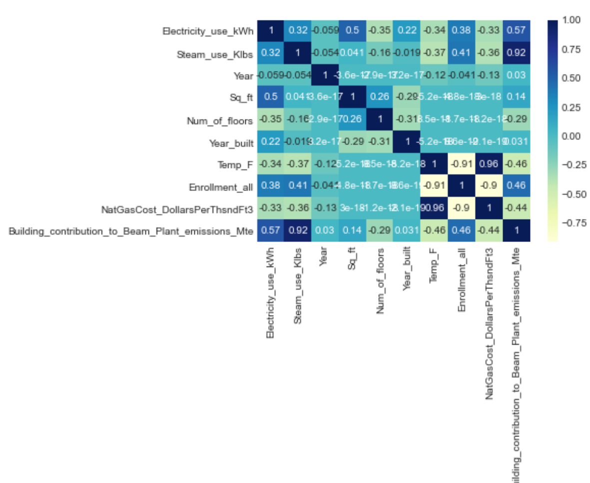

Figure 1: Correlation matrix for the reduced data set.

These visuals allowed us to investigate the relationship between our response (Building's fraction of Beam Plant emissions, Mte) and the individual variables of our reduced data set. After this initial exploration, we performed further feature selection by means of the mutual information measure and an F-test (see Figure 2 below). Unlike the F-test, the mutual information method assigned significance to additional independent variables besides steam.

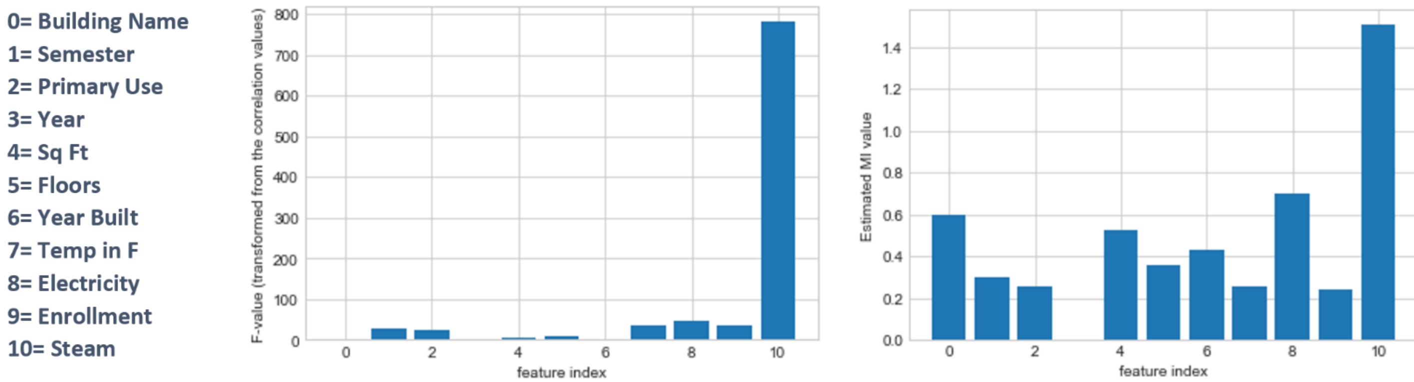

Figure 2: F-test (left) and mutual information (right) results. 30% of the available data was used to train, while 70% was used for the test. The labels for the horizontal axes are explained by the column on the far left.

Both the F-test and the mutual information measure revealed that Year was not a useful predictor, therefore, the "full model" of our data set, which we used for our random forest and neural network models, does not include Year.

II.II Multiple Linear Regression

Considering the number of data points (n) available for each building in our data set, it made sense to incorporate a more rigid model into our project, especially as a jumping off point. A pair of multiple linear regression models were built for the purposes of inference and/or quantitative prediction. For both regression models, the average temperature, enrollment, cost of natural gas in Michigan, and the electricity and steam used at the building level were selected as the input variables by which our output variable (Building's fraction of Beam Plant emissions, Mte) was predicted.

Correlation values:		Selected?
Electricity_use_kWh	0.567537	Yes
Steam_use_Klbs	0.917963	Yes
Year	0.029604	No
Sq_ft	0.143954	No
Num_of_floors	0.288821	No
Year_built	0.030963	No
Temp_F	0.459350	Yes
Enrollment_all	0.462341	Yes
NatGasCost_DollarsPerThsndFt3	0.442200	Yes

Table 1: Variables selected according to correlation values for the multiple linear regression (MLR) models.

From here on, the model that we refer to as the "reduced" data set is the set that has had low-correlation value variables eliminated. The first multiple linear regression model was trained and then tested using all the available data for these variables. Our second model was built after removing select outliers from the same data set, bringing the total number of rows down from 196 to 189. It should be noted that for these linear regression algorithms, as well as for every algorithm incorporated into this project, 80% of the available data was used to train the model and the remaining 20% was used for performance testing. Mathematically, both multiple linear regression models took the following form:

Output = Intercept + (A × Enrollment) + (B × Temperature) + (C × Electricity) + (D × Steam) + (E × Cost)

	MLR Model 1	MLR Model 2
Enrollment Coefficient (A)	-1.09083402e-05	-8.63633973e-06
Temperature Coefficient (B)	-4.84647858e-03	-1.25620106e-03
Electricity Use Coefficient (C)	5.70806043e-07	3.80832011e-07

Steam Use Coefficient (D)	6.93930953e-05	1.01553950e-04
Natural Gas Cost Coefficient (E)	-3.05471033e-02	-2.60977986e-02
Test RMSE	0.09	0.05
Test R^2	0.93	0.97

Table 2: Comparison of both coefficients and test performance for the two multiple linear regression (MLR) models.

By removing select outlying data points, we see a notable increase in multiple linear regression performance.

II.III Random Forest Models

While linear regression was a sensible starting point, we were interested in exploring the applicability of other algorithms that we have learned about during the semester. Following this curiosity, the next step we took was to develop a random forest model. Four total random forest models were observed; two models used the full data set and the other two only used the reduced data set. For each data set applied to this model, two cases were considered: 1., when the model has a maximum depth of 4; and, 2., when the model's maximum depth is not specified. In all four models, the number of estimators was set to equal 100 and the random state was specified as 0. The models' respective performances are compared in Table 3 below.

	Maximum Depth = 4	Maximum Depth Unspecified
Reduced Data RMSE	0.05	0.02
Full Data RMSE	0.031	0.017
Reduced Data R^2	0.97	0.996
Full Data R^2	0.99	0.997

Table 3: Comparison of test performance between the four random forest models.

While limiting the maximum depth of the random forest model may be beneficial for the sake of saving on computational power, it does seem to hamper the model's accuracy in this scenario. Also, Table 3 above shows an increase in random forest model test performance when more of our variables are incorporated (in comparing the full versus reduced model).

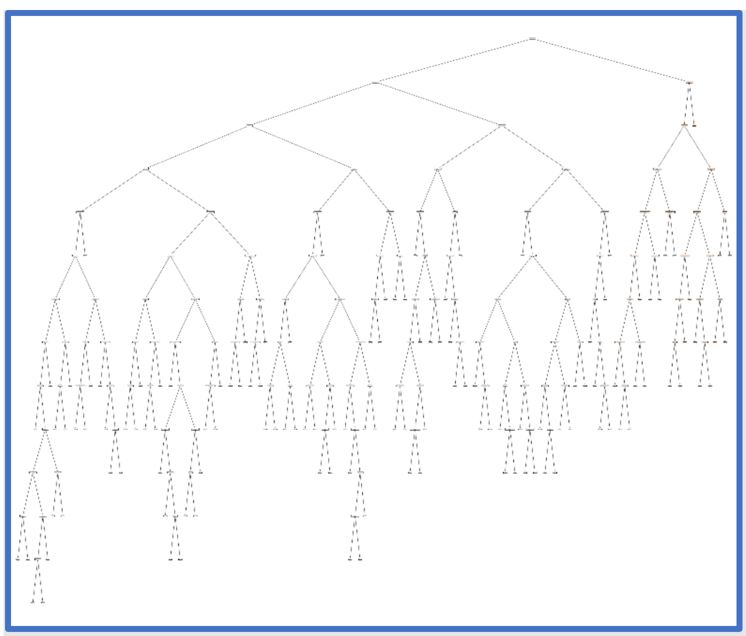

Figure 3: Random forest model, reduced data set (max depth unspecified).

II.IV Neural Networks

Finally, our group was interested in seeing how applicable a neural network may or may not have been to our data. Sixteen neural networks were observed. Like the random forest models, half of the neural networks were trained on the reduced data set while the other half utilized the full data set (again, with calendar year being factored out). Thirty-five rows of the original data set were randomly selected as the test set and set aside in a separate excel sheet. This reduced the data set to 161 rows for the training and validation portion. To prevent data leakage, normalization[xxxiii] was applied after splitting the data into test and training sets. Then, the input, output and hidden layers were calculated for the full and reduced models using the equations below:

Output Layer:

1

Note: Output layer is always equal to 1, because there is only one response variable.

Input Layer:

P+1

Hidden Layer:

$$\frac{\text{Training Data Samples}}{Factor} * (\text{input layer} + \text{output layer})$$

Factor:

Takes on a chosen value between 1-10. The purpose of this value is to prevent overfitting.

Note: for simplicity, all 16 models used a factor of 1.

Training data Samples:

N * .80

The table below (Table 4) shows the calculations and resulting values for both the reduced and full data sets:

Model	P	N (Validation/ Training)	Hidden Layer	Input Neurons	Output Neurons	Factor	Training Data Samples
Full Model (-year)	11	161	$\frac{129}{1}$ * (12+ 1) = 1677	11+1 = 12	1	1	161*.80 = 129
Reduced Model	5	161	$\frac{129}{1}$ * (6+ 1) = 907	5+1 = 6	1	1	161*.80 = 129

Table 4: Neural network parameters for both the full and reduced data set models.

When working with neural networks, it is important to understand the trade-off between batch-size and number of epochs, as well as how the activation function can influence model performance.

We chose two activation functions to run trials on, Exponential Linear Units (ELU) and Rectified Linear Units (ReLU). We found that ELU gave lower test RMSE scores in both the reduced and full models. We also discovered that, as epoch size increased, test RMSE decreased. Additionally, as batch size simultaneously increased up to a value of 85, the test RMSE decreased. After the batch size increased past 85, the test RMSE increased. We therefore found that our optimal parameters for both models were 100 and 85 for epoch and batch size, respectively. The graph below depicts this trend.

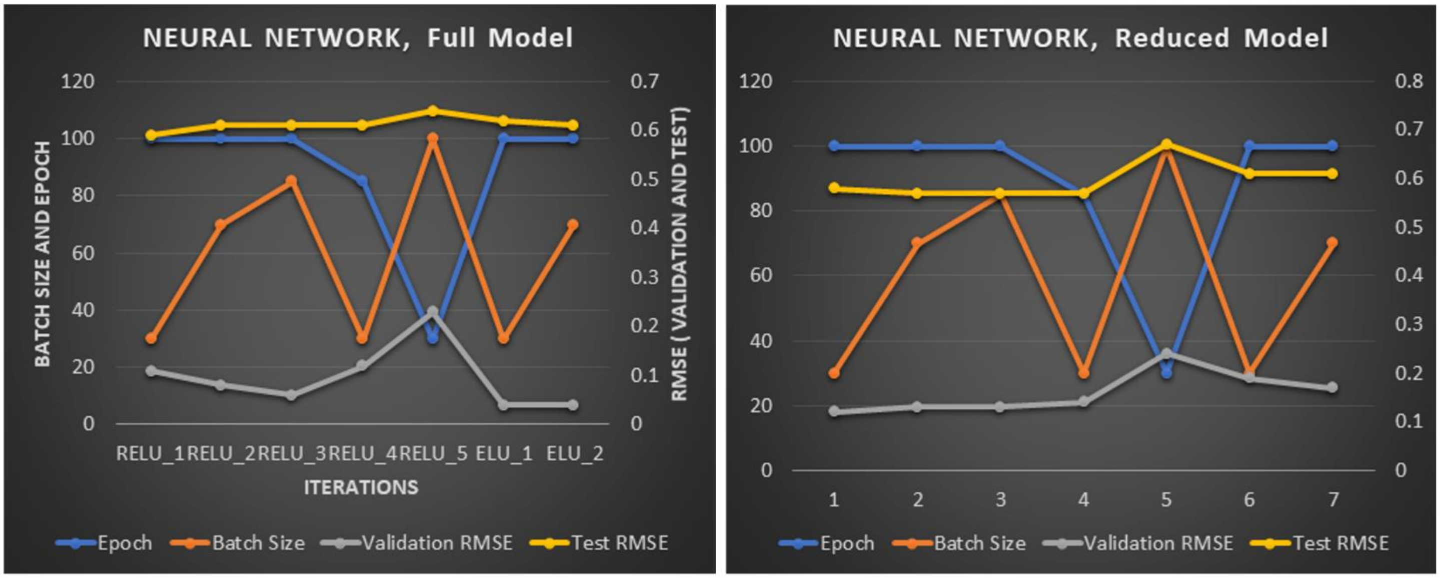

Figure 4: Neural network parameters and model performance.

The two best models for neural network are shown below, in Figures 6 and 7. The MSE score is used as the loss function on the y-axis.

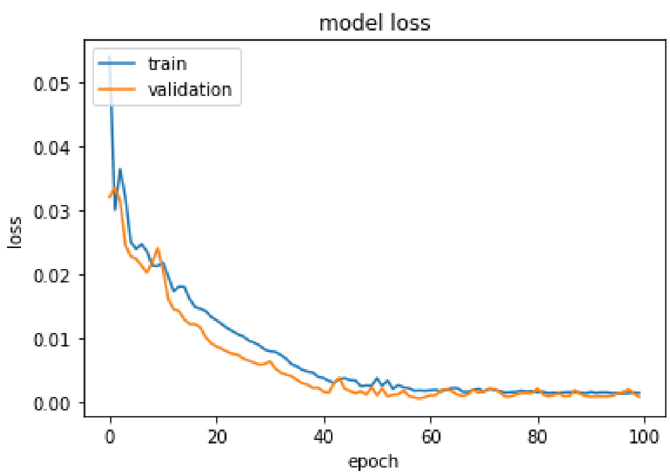

Figure 5: Neural network, full model (- year): epoch = 100; batch = 85; activation = ELU

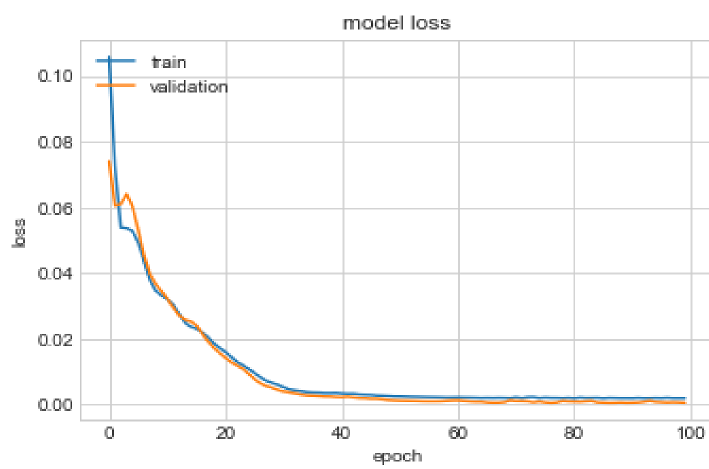

Figure 6: Neural network, full reduced model (- year): epoch = 100; batch = 85; activation = ELU

While the neural network models all produced lower validation RMSE, we found that the test RMSE was on average 0.51 RMSE units higher than the validation set. This indicates that our neural network models were either overfitting, or the relationship between our variables were closer to true linear.

III. Key Findings

In exploring the correlation between the variables of our data set, we found that calendar year had a negligible influence on our output variable. Omitting outlying data points increased the accuracy and predictive power of our multiple linear regression model. Since the neural network test RMSE was significantly higher than the validation/training RMSE, our neural networks were either overfitting, or, our variables had closer to a true linear relationship. Our random forest models exhibited equivalent if not superior test data performance compared to the multiple linear regression models. Despite this, multiple linear regression yielded higher predictive accuracy over random forest models and neural networks when tested on the 2020 data point. This corroborates the usefulness of rigid models in scenarios when the number of data points (n) is relatively small compared to the number of predictors (p), which is the case in this project.

Model	R^2	RMSE
LR #1 (reduced P & outliers NOT removed)	0.93	0.09
LR #2 (reduced P & outliers removed)	0.97	0.05
NN 1 (full, - year)	-	0.58
NN 2 (reduced P)	-	0.53
RF 1 (full, - year & max depth = 4)	0.99	0.03
RF 2 (full, - year & max depth = unspecified)	0.99	0.01
RF 3 (Reduced P, - year & max depth = 4)	0.97	0.05
RF 4 (Reduced P, - year & max depth = unspecified)	0.99	0.02

Table 5: Model performance compared. Random forest on the full data set (minus calendar year) with maximum depth unspecified performed the best in terms of R^2 and RMSE.

Actual value CO2 MTE, Bernhard Center, Jan 2020 = 0.319214		
Top Three Models	**Results**	**Difference**
LR #2	0.31548995	0.003
Random Forest (full model -year) [Max_depth not specified]	0.376747188	-0.057
Random Forest (Reduced Model) [[Max_depth not specified]	0.37616963	-0.056

Table 6: Model prediction compared. The multiple linear regression model with outliers removed performed best in terms of prediction accuracy on January 2020 data for the Bernhard Center.

IV. Conclusion

Building the neural networks was an informative process, and one that most if not all of us are likely to use again in the future. While it was the most complex of our model types, it was also the least accurate of the three model types that we explored. The random forest models performed well on the test data but did not predict the 2020 data as accurately as the multiple linear regression model. By increasing the number of predictors in our random forest model and by leaving the maximum depth unspecified, we were able to improve the test performance.

Ultimately, we hope that even a small part of our work here contributes to progress towards a carbon price, or, at least better energy bill infrastructure and CO2 emissions accounting for budget units here on campus.

In the future, an easy additional test of our finding that the multiple linear regression model is the best at predicting actual values (from 2020) is to test additional actual 2020 values on this model.

To further assist in the efforts to understand CO2 emissions accountability and electricity billing amongst college buildings, we would first want to identify a few budget units under the new, decentralized budget model for which metering for (at least) steam and electricity use is complete. Next, we would want to add additional building-relevant variables such as insulation type, date of insulation installment, window type, date of window installment, roof material, date of roof material installment and outer wall material, as well as any additional relevant factors. Finally, we would want to use the model to predict emissions by semester or financial quarter (whichever period WMU has chosen for energy use billing purposes)—particularly for new buildings like the new student center.

Further, we are aware that 15-20% of the energy coming out of the Beam plant to campus buildings is electricity that is purchased from Consumers, which is generated by a mix of sources. For our emissions calculations to be 100% accurate, we should factor in the from Consumers.

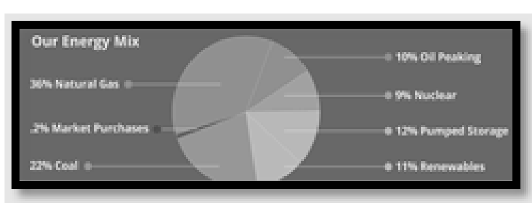

Figure 8: Consumers Energy general energy mix.

V. References

Brownlee, K. (2019, October 16), Machine Learning Mastery: Information Gain and Mutual Information for Machine Learning. Retrieved from: https://machinelearningmastery.com/information-gain-and-mutual-information/

Consumers Energy. (2021). Our energy mix: Consumers energy. Retrieved from https://www.consumersenergy.com/community/sustainability/energy-mix

EPA. (2019). 2019 GHG emissions form large facilities. Retrieved from https://ghgdata.epa.gov/ghgp/main.do#/facility/

Galarnyk, M. (2021, February 03). Visualizing decision trees with Python (SCIKIT-LEARN, Graphviz, Matplotlib). Retrieved from https://towardsdatascience.com/visualizing-decision-trees-with-python-scikit-learn-graphviz-matplotlib-1c50b4aa68dc

Glen, S. (2021). StatisticsHowTo: Elementary Statistics for the rest of us! Retrieved from: https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/

Gordon, M. (2020, May 15). Towards Data Science: Regression-based neural networks: Predicting Average Daily Rates for Hotels. Retrieved from https://towardsdatascience.com/regression-based-neural-networks-with-tensorflow-v2-0-predicting-average-daily-rates-e20fffa7ac9a

Grogan, M. (2020, November 07). One-step predictions with LSTM: Forecasting hotel revenues. Retrieved from https://towardsdatascience.com/one-step-predictions-with-lstm-forecasting-hotel-revenues-c9ef0d3ef2df

Grogan, M. (2020, October 14). Regression-based neural networks with Tensorflow v2.0: Predicting average daily rates. Retrieved from https://towardsdatascience.com/regression-based-neural-networks-with-tensorflow-v2-0-predicting-average-daily-rates-e20fffa7ac9a

Jarvis, G. (2021, April 9). [Beam power plant standard conversion of steam usage to BTu as 1lb of steam to 1,060 BTu's]. Unpublished raw data. [This is the Beam power plant/industry standard, as far as we understand. In general, MMBtu = 1 Million Btu = dekatherm = one thousand thousand British thermal units. Then, according to our standard, 1,000 lbs = 1.060 MMBTus. More high-pressure steam system conversions have different rates, such as at the U of Illinois: https://icap.sustainability.illinois.edu/project-update/klbs-steam-mmbtu-conversion-factor-campus. Other, general information on steam to BTu conversion is at: https://sciencing.com/convert-steam-btu-8495235.html]

Jarvis, G., & Janson, A. Discussion of historical WMU energy use across campus. [Telephone interview series]. (2021, March - April).

Jarvis, G., & Janson, A. Discussion of historical WMU power plant data components. [Personal interview]. (2021, March 19).

Jarvis, G., & WMU Facilities Management. (2021, March 18). [Annual historical balance condenser condensate]. Unpublished raw data.

Jarvis, G., & WMU Facilities Management. (2021, April 11). [Monthly electricity output: Beam Power Plant]. Unpublished raw data.

Jarvis, G., & WMU Facilities Management. (2021, April 11). [Monthly steam output: Beam Power Plant]. Unpublished raw data.

Jarvis, G., & WMU Facilities Management. (2021, April 11). [Monthly natural gas use: Beam Power Plant]. Unpublished raw data.

Koehrsen, W. (2018, January 08). Improving the random forest in python part 1. Retrieved from https://towardsdatascience.com/improving-random-forest-in-python-part-1-893916666cd

Koehrsen, W. (2018, January 17). Random forest in python. Retrieved from https://towardsdatascience.com/random-forest-in-python-24d0893d51c0

Mayo, M. (2017, september). KDnuggets™: Neural Network Foundations, Explained: Activation Function. Retrieved from: https://www.kdnuggets.com/2017/09/neural-network-foundations-explained-activation-function.html

Mwiti, D. (2021, April 09). Random forest regression: When does it fail and why? Retrieved from https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why

Pung, K., Ward, J., & Janson, A. (2021, April 09). [Email discussions with Keith Pung and Joel Ward]. Unpublished raw data.

Ramos, L. (2021). Calculating mean, median and mode in python. Retrieved from https://stackabuse.com/calculating-mean-median-and-mode-in-python/

Strazdas, P., & Janson, A. Discussion of ML project objectives and parameters [Telephone interview]. (2021, April 12).

Time and Date AS. (2020). Climate & Weather averages in Kalamazoo, Michigan, USA. Retrieved from https://www.timeanddate.com/weather/usa/kalamazoo/climate

United States Environmental Protection Agency. (2018, October 15). Greenhouse gas equivalencies calculator. Retrieved from https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator [xxxiv]

United States Environmental Protection Agency. (2021, February 09). Greenhouse gases equivalencies calculator - calculations and references. Retrieved from https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references

UnitJuggler. (2021). Convert kilowatt hours to million Btu: Energy converter. Retrieved April 9, 2020, from https://www.unitjuggler.com/convert-energy-from-kWh-to-MMBtu.html [Electricity use at 1 kWh = .00340951 MMBTu.]

US Energy Information Administration. (2021, March 31). Michigan price of natural gas sold to commercial consumers (dollars per thousand cubic feet). Retrieved from https://www.eia.gov/dnav/ng/hist/n3020mi3m.htm

US EPA. (2021, March 11). Greenhouse gases equivalencies calculator: Calculations and references. Retrieved from https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references [The conversion rate cited by the EPA is .0548. "The average carbon dioxide coefficient of natural gas is 0.0548 kg CO2 per cubic foot. The fraction oxidized to CO2 is 100 percent." The EPA gas equivalencies calculator, at https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator, rounds to the nearest hundredth. Considering these facts, we used .0549 as our equivalency.]

Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

Western Michigan University. (2019, April 04). Strategic resource management. Retrieved from https://wmich.edu/budgetmodel

Western Michigan University. (2021). Institutional Research enrollment. Retrieved from https://wmich.edu/institutionalresearch/reporting/enrollment

Western Michigan University Facilities Management. (2021). WMU Building List. Retrieved from https://www.fm.wmich.edu/ap/bldg/buildingList?sort=code

Yale University. (2016, October 10). Yale University's carbon charge: Preliminary results from learning by doing. Retrieved from https://carbon.yale.edu/sites/default/files/files/Carbon_Charge_Pilot_Report_20161010.pdf

---

[i] EPA. (2019). 2019 GHG emissions form large facilities. Retrieved from https://ghgdata.epa.gov/ghgp/main.do#/facility/

[ii] Western Michigan University. (2019, April 04). Strategic resource management. Retrieved from https://wmich.edu/budgetmodel

[iii] Yale University. (2016, October 10). Yale University's carbon charge: Preliminary results from learning by doing. Retrieved from https://carbon.yale.edu/sites/default/files/files/Carbon_Charge_Pilot_Report_20161010.pdf

[iv] Strazdas, P., & Janson, A. Discussion of ML project objectives and parameters [Telephone interview]. (2021, April 12).

[v] Strazdas, P., & Janson, A. Discussion of ML project objectives and parameters [Telephone interview]. (2021, April 12).

[vi] Jarvis, G., & Janson, A. Discussion of historical WMU energy use across campus. [Telephone interview series]. (2021, March - April).

[vii] Jarvis, G., & WMU Facilities Management. (2021, March 18). [Annual historical balance condenser condensate]. Unpublished raw data.

[viii] Strazdas, P., & Janson, A. Discussion of ML project objectives and parameters [Telephone interview]. (2021, April 12).

[ix] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[x] The "Building Offline" status symbols should be updated for some of the buildings in the list of those included in this study at https://www.fm.wmich.edu/ap/bldg/buildingList?sort=code.

[xi] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[xii] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[xiii] UnitJuggler. (2021). Convert kilowatt hours to million Btu: Energy converter. Retrieved April 9, 2020, from https://www.unitjuggler.com/convert-energy-from-kWh-to-MMBtu.html [Electricity use at 1 kWh = .00340951 MMBTu.]

[xiv] Jarvis, G. (2021, April 9). [Estimated conversion of steam usage to BTu as 1lb of steam to 1,060 BTu's]. Unpublished raw data. [This is the Beam power plant/industry standard, as far as we understand. In general, MMBtu = 1 Million Btu = dekatherm = one thousand thousand British thermal units. Then, according to our standard, 1,000 lbs = 1.060 MMBTus. More high-pressure steam system conversions have different rates, such as at the U of Illinois: https://icap.sustainability.illinois.edu/project-update/klbs-steam-mmbtu-conversion-factor-campus. Other, general information on steam to BTu conversion is at: https://sciencing.com/convert-steam-btu-8495235.html]

[xv] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[xvi] US EPA. (2021, March 11). Greenhouse gases equivalencies calculator: Calculations and references. Retrieved from https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references [The conversion rate cited by the EPA is .0548. "The average carbon dioxide coefficient of natural gas is 0.0548 kg CO2 per cubic foot. The fraction oxidized to CO2 is 100 percent." The EPA gas equivalencies calculator, at https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator, rounds to the nearest hundredth. Considering these facts, we used .0549 as our equivalency.]

[xvii] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[xviii] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[xix] WMU Facilities Management. (2021). WMU Building List. Retrieved from https://www.fm.wmich.edu/ap/bldg/buildingList?sort=code

[xx] WMU Facilities Management. (2021). WMU Building List. Retrieved from https://www.fm.wmich.edu/ap/bldg/buildingList?sort=code

[xxi] WMU Facilities Management. (2021). WMU Building List. Retrieved from https://www.fm.wmich.edu/ap/bldg/buildingList?sort=code

[xxii] WMU Facilities Management. (2021). WMU Building List. Retrieved from https://www.fm.wmich.edu/ap/bldg/buildingList?sort=code

[xxiii] Time and Date AS. (2020). Climate & Weather averages in Kalamazoo, Michigan, USA. Retrieved from https://www.timeanddate.com/weather/usa/kalamazoo/climate

[xxiv] Western Michigan University. (2021). Institutional Research Enrollment. Retrieved from https://wmich.edu/institutionalresearch/reporting/enrollment

[xxv] US Energy Information Administration. (2021, March 31). Michigan price of natural gas sold to commercial consumers (dollars per thousand cubic feet). Retrieved from https://www.eia.gov/dnav/ng/hist/n3020mi3m.htm

[xxvi] Ward, J., & WMU Facilities Management. (2021, April 11). [Student Affairs Utility Consumption Report]. Unpublished raw data.

[xxvii] Jarvis, G. (2021, April 9). [Beam power plant standard conversion of steam usage to BTu as 1lb of steam to 1,060 BTu's]. Unpublished raw data. [This is the Beam power plant/industry standard, as far as we understand. In general, MMBtu = 1 Million Btu = dekatherm = one thousand thousand British thermal units. Then, according to our standard, 1,000 lbs = 1.060 MMBTus. More high-pressure steam system conversions have different rates, such as at the U of Illinois: https://icap.sustainability.illinois.edu/project-update/klbs-steam-mmbtu-conversion-factor-campus. Other, general information on steam to BTu conversion is at: https://sciencing.com/convert-steam-btu-8495235.html]

[xxviii] UnitJuggler. (2021). Convert kilowatt hours to million Btu: Energy converter. Retrieved April 9, 2020, from https://www.unitjuggler.com/convert-energy-from-kWh-to-MMBtu.html [Electricity use at 1 kWh = .00340951 MMBTu.]

[xxix] Jarvis, G., & WMU Facilities Management. (2021, April 11). [Monthly electricity output: Beam Power Plant]. Unpublished raw data.

[xxx] Jarvis, G., & WMU Facilities Management. (2021, April 11). [Monthly steam output: Beam Power Plant]. Unpublished raw data.

[xxxi] Jarvis, G., & WMU Facilities Management. (2021, April 11). [Monthly natural gas use: Beam Power Plant]. Unpublished raw data.

[xxxii] US EPA. (2021, March 11). Greenhouse gases equivalencies calculator: Calculations and references. Retrieved from https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references [The conversion rate cited by the EPA is .0548. "The average carbon dioxide coefficient of natural gas is 0.0548 kg CO2 per cubic foot. The fraction oxidized to CO2 is 100 percent." The EPA gas equivalencies calculator, at https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator, rounds to the nearest hundredth. Considering these facts, we used .0549 as our equivalency.]

[xxxiii] Grogan, M. (2020, October 14). Regression-based neural networks with Tensorflow v2.0: Predicting average daily rates. Retrieved from https://towardsdatascience.com/regression-based-neural-networks-with-tensorflow-v2-0-predicting-average-daily-rates-e20fffa7ac9a

[xxxiv] United States Environmental Protection Agency. (2018, October 15). Greenhouse gas equivalencies calculator. Retrieved from https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator