ARTIFICIAL INTELLIGENCE IN SMART GRID FOR ENERGY MANAGEMENT

Chibuzo Valentine Nwadike

Electrical and Computer Engineering & Illinois Institute of Technology, Chicago, Illinois, 60616, United States

eit 2023
at Lewis University

Goal of the Research

The project aims to provide insights into the efficacy of using temperature data from specific stations for predicting load values. And to help inform utility companies in their decision-making processes related to load forecasting and resource allocation, leading to more efficient and effective energy management strategies.

Scope

Categories and Subject Descriptors
Machine Learning, Forecasting
Keywords
Data, Load, Stations, Correlation, Regression, Mode, Hyperparameters.

Abstract

This project focuses on analysing load and temperature data collected from utility companies in the US, consisting of 20 load zones with varying patterns of hourly load values and 11 temperature stations with distinct locations. The objective is to identify patterns and correlations between temperature data and load values for each zone and develop a predictive model for load values using machine learning algorithms.

A thorough data exploration was conducted to examine potential correlations between temperature stations and load values in each zone. In cases where strong correlations were found, the temperature data from the correlated station was utilized to predict load values in the corresponding zones. However, in instances where strong correlations were not identified, a method was devised to select temperature data from a station and incorporate it into machine learning algorithms for predicting load values in each load zone. The methodology for selecting temperature data for load prediction involves various factors such as geographical proximity, climatic similarity, historical data analysis, and statistical measures. Machine learning algorithms are applied to develop predictive models using the selected temperature data.

The results of this project will contribute to a better understanding of the relationship between temperature and load values in different load zones and provide insights into the efficacy of using temperature data from specific stations for predicting load values. This research helps to inform utility companies in their decision-making processes related to load forecasting and resource allocation, ultimately leading to more efficient and effective energy management strategies.

Problem Description

The problem addressed in this project is related to load forecasting and energy management in the utility industry. Specifically, the project aims to analyze the relationship between temperature and load values in different load zones and develop a predictive model for load values using machine learning algorithms.

The challenge is that load values can vary significantly based on factors such as time of day, day of the week, and weather conditions. Temperature is one of the key weather variables that can impact load values, as temperature changes can influence the use of heating and cooling systems in residential and commercial buildings. Therefore, identifying patterns and correlations between temperature data and load values can help utility companies make more accurate load forecasts and allocate resources more effectively.

Proposed Method

- ➢Acquired dataset
- ➢Preprocess the dataset
- ➢Find Patterns and correlations between temperature and load values and perform mapping.
- ➢Build six models using MLPRegressor, Linear Regression, Decision Tress, Random Forset, KNN Regression, Xgboost.
- ➢Compare the test scores of each model to find the model with high prediction score.
- ➢Use the best model to predict with a high accuracy, the load values for the week of June 1-7, 2008 for each load zone, given the temperature and the historical load data.

Experimental Results

		zone	station	corr
1	0	zone_1	station_1	0.8834255626564089
2	1	zone_2	station_7	0.8953481481234017
3	2	zone_3	station_9	0.747905238425857
4	3	zone_4	station_11	0.8901578678729276
5	4	zone_5	station_10	0.8737992309525794
6	5	zone_6	station_9	0.7479859609436017
7	6	zone_7	station_9	0.7653480625386652
8	7	zone_8	station_10	0.876036767568502
9	8	zone_9	station_4	0.7924677836751881
10	9	zone_10	station_11	0.8780894404851813
11	10	zone_11	station_7	0.5922642640881565
12	11	zone_12	station_7	0.8854012907700866
13	12	zone_13	station_1	0.9039429503215645
14	13	zone_14	station_8	0.7965535884798502
15	14	zone_15	station_10	0.825658539548006
16	15	zone_16	station_7	0.8685641842281417
17	16	zone_17	station_9	0.7481895210035447
18	17	zone_18	station_7	0.7897704497519189
19	18	zone_19	station_6	0.1356449258524945
20	19	zone_20	station_9	0.8757040228300434

Figure 1: mapping of a temperature station for each load zone

Decision Tree Model

		zone_id	year	month	day	hour	y_pred_load	y_true_load	relative percentage er...
1	0	13	2005	5	17	h14	153803.0	41765	268.26
2	1	10	2005	5	23	h14	767760.0	212158	261.88
3	2	15	2005	6	21	h5	79046.0	22129	257.2
4	3	10	2005	5	23	h15	748891.0	211534	254.03
5	4	10	2005	5	23	h16	751198.0	212343	253.77
6	5	5	2005	5	23	h15	5536202.0	1569412	252.76
7	6	13	2005	5	17	h15	148733.0	42492	250.03
8	7	5	2005	5	23	h16	5584884.0	1608618	247.19
9	8	5	2005	5	7	h17	5688421.0	1667519	241.13
10	9	5	2005	5	23	h17	5688421.0	1689922	236.61

Figure 3: Top ten prediction errors

		zone	station	fit_train_time	test_score
1	0	1	1	0.0646	0.66
2	1	2	7	0.0562	0.71
3	2	3	9	0.065	0.43
4	3	4	11	0.0609	0.7
5	4	5	10	0.0545	0.75
6	5	6	9	0.0673	0.43
7	6	7	9	0.0687	0.5
8	7	8	10	0.0598	0.76
9	8	9	4	0.0617	0.56
10	9	10	11	0.0642	0.71
11	10	11	7	0.0552	0.3
12	11	12	7	0.0554	0.65
13	12	13	1	0.0637	0.68
14	13	14	8	0.0655	0.6
15	14	15	10	0.0699	0.54
16	15	16	7	0.0631	0.68
17	16	17	9	0.0665	0.43
18	17	18	7	0.0596	0.56
19	18	20	9	0.0674	0.66

Figure 4: Train time and test scores

KNN Regression Model

		zone_id	year	month	day	hour	y_pred_load	y_true_load	relative percentage er...
1	0	13	2005	6	2	h17	117590.0	32473	262.12
2	1	10	2005	5	23	h16	750945.0	212343	253.65
3	2	13	2005	5	17	h14	147630.0	41765	253.48
4	3	10	2005	5	23	h15	736254.0	211534	248.05
5	4	13	2005	6	2	h15	123165.0	35613	245.84
6	5	10	2005	5	23	h14	729980.0	212158	244.07
7	6	13	2006	5	24	h14	136711.0	39916	242.5
8	7	13	2005	5	17	h15	145394.0	42492	242.17
9	8	15	2005	6	21	h5	75351.0	22129	240.51
10	9	13	2005	5	7	h19	177004.0	52482	237.27

Figure 5: Top ten prediction errors

		zone	station	fit_train_time	test_score
1	0	1	1	1.0041	0.74
2	1	2	7	0.993	0.79
3	2	3	9	1.0042	0.55
4	3	4	11	0.9964	0.79
5	4	5	10	0.9959	0.82
6	5	6	9	1.0406	0.55
7	6	7	9	1.0206	0.58
8	7	8	10	1.095	0.81
9	8	9	4	0.9959	0.65
10	9	10	11	0.9978	0.79
11	10	11	7	0.9949	0.35
12	11	12	7	0.9636	0.73
13	12	13	1	0.9762	0.77
14	13	14	8	1.2495	0.69
15	14	15	10	0.9994	0.64
16	15	16	7	0.9933	0.74
17	16	17	9	0.9838	0.55
18	17	18	7	0.9937	0.63
19	18	20	9	0.9939	0.73

Figure 6: Train time and test scores

MLPRegressor Model

		zone_id	year	month	day	hour	y_pred_load	y_true_load	relative percentage er...
1	0	15	2005	6	21	h5	82834.0	22129	274.33
2	1	13	2005	6	2	h17	112996.0	32473	247.97
3	2	13	2005	6	2	h15	108667.0	35613	205.13
4	3	15	2005	6	21	h4	81803.0	27201	200.74
5	4	13	2005	6	2	h14	108958.0	37139	193.38
6	5	15	2005	6	21	h2	86841.0	29624	193.14
7	6	15	2005	6	21	h3	83307.0	28611	191.17
8	7	13	2005	6	2	h16	109883.0	37813	190.59
9	8	13	2006	5	16	h16	113665.0	40306	182.01
10	9	13	2005	5	5	h14	112936.0	40149	181.29

Figure 7: Top ten prediction errors

		zone	station	fit_train_time	train_score	test_score
1	0	1	1	0.8451	0.55	0.54
2	1	2	7	0.7847	0.47	0.47
3	2	3	9	0.9102	0.43	0.43
4	3	4	11	0.9726	0.4	0.39
5	4	5	10	0.8621	0.46	0.46
6	5	6	9	0.7195	0.1	0.13
7	6	7	9	0.7192	-0.06	-0.03
8	7	8	10	0.7332	0.48	0.47
9	8	9	4	0.6962	0.43	0.42
10	9	10	11	0.6979	0.45	0.45
11	10	12	7	0.7881	0.5	0.49
12	11	13	1	0.6148	0.45	0.45
13	12	14	8	0.6666	0.31	0.31
14	13	15	10	0.6358	0.49	0.49
15	14	16	7	0.6999	0.45	0.45
16	15	17	9	0.646	0.42	0.41
17	16	18	7	0.5968	0.33	0.33
18	17	20	9	0.6394	0.42	0.42

Figure 8: Train time and test scores

Linear Regression Model

		zone_id	year	month	day	hour	y_pred_load	y_true_load	relative percentage er...
1	0	13	2005	5	7	h16	165822.0	45331	265.8
2	1	13	2005	5	7	h17	168074.0	46422	262.06
3	2	13	2005	6	2	h17	115626.0	32473	256.07
4	3	13	2005	5	17	h14	147590.0	41765	253.38
5	4	15	2005	6	21	h5	77307.0	22129	249.35
6	5	13	2005	5	7	h19	179933.0	52482	242.85
7	6	13	2005	5	7	h15	164884.0	48266	241.61
8	7	13	2005	5	17	h15	144823.0	42492	240.82
9	8	13	2005	6	2	h15	120699.0	35613	238.92
10	9	13	2005	5	5	h14	135089.0	40149	236.47

Figure 9: Top ten prediction errors

		zone	station	fit_train_time	train_score	test_score
1	0	1	1	0.0967	0.72	0.66
2	1	2	7	0.0726	0.7	0.65
3	2	3	9	0.0667	0.55	0.49
4	3	4	11	0.0777	0.7	0.66
5	4	5	10	0.0762	0.73	0.68
6	5	6	9	0.0768	0.55	0.49
7	6	7	9	0.0605	0.57	0.51
8	7	8	10	0.0647	0.74	0.69
9	8	9	4	0.0623	0.58	0.48
10	9	10	11	0.0715	0.7	0.66
11	10	12	7	0.0645	0.69	0.64
12	11	13	1	0.0779	0.73	0.68
13	12	14	8	0.0771	0.55	0.47
14	13	15	10	0.079	0.65	0.6
15	14	16	7	0.0548	0.66	0.6
16	15	17	9	0.0683	0.55	0.49
17	16	18	7	0.0752	0.52	0.44
18	17	20	9	0.0774	0.65	0.6

Figure 10: Train time and test scores

Random Forest Model

		zone_id	year	month	day	hour	y_pred_load	y_true_load	relative percentage er...
1	0	13	2005	5	17	h14	146577.0	41765	250.96
2	1	10	2005	5	23	h16	742612.0	212343	249.72
3	2	13	2005	6	2	h17	113350.0	32473	249.06
4	3	5	2005	5	7	h16	5803890.0	1664848	248.61
5	4	5	2005	5	7	h17	5777027.0	1667519	246.44
6	5	10	2005	5	23	h15	731180.0	211534	245.66
7	6	10	2005	5	23	h14	726100.0	212158	242.24
8	7	5	2005	5	23	h15	5368453.0	1569412	242.07
9	8	15	2005	6	21	h5	75589.0	22129	241.58
10	9	13	2005	5	17	h15	144340.0	42492	239.69

Figure 11: Top ten prediction errors

		zone	station	fit_train_time	test_score
1	0	1	1	4.0232	0.76
2	1	2	7	4.8284	0.81
3	2	3	9	4.4655	0.58
4	3	4	11	4.194	0.81
5	4	5	10	4.4767	0.84
6	5	6	9	4.2462	0.58
7	6	7	9	4.0888	0.61
8	7	8	10	4.2109	0.83
9	8	9	4	4.0105	0.67
10	9	10	11	4.1004	0.81
11	10	12	7	4.1841	0.76
12	11	13	1	3.991	0.79
13	12	14	8	4.0365	0.73
14	13	15	10	4.7168	0.68
15	14	16	7	4.3018	0.76
16	15	17	9	4.1748	0.59
17	16	18	7	5.3514	0.67
18	17	20	9	4.2847	0.75

Figure 12: Train time and test scores

Xgboost

		zone_id	year	month	day	hour	y_pred_load	y_true_load	relative percentage er...
1	0	13	2005	6	2	h15	139203.0	35613	290.88
2	1	13	2005	5	17	h14	156162.0	41765	273.91
3	2	13	2005	6	2	h17	120433.0	32473	270.87
4	3	10	2005	5	23	h16	778754.0	212343	266.74
5	4	10	2005	5	23	h15	749972.0	211534	254.54
6	5	13	2005	5	17	h15	149758.0	42492	252.44
7	6	13	2006	5	24	h14	139206.0	39916	248.75
8	7	5	2005	5	7	h16	5802114.0	1664848	248.51
9	8	10	2005	5	23	h17	784396.0	225173	248.35
10	9	10	2005	6	2	h15	649844.0	186933	247.63

Figure 13: Top ten prediction errors

		zone	station	fit_train_time	test_score
1	0	1	1	10.9392	0.72
2	1	2	7	11.4974	0.78
3	2	3	9	11.3154	0.49
4	3	4	11	12.1877	0.78
5	4	5	10	12.0029	0.81
6	5	6	9	15.3059	0.49
7	6	7	9	22.6034	0.52
8	7	8	10	15.9119	0.8
9	8	9	4	11.9133	0.61
10	9	10	11	12.468	0.79
11	10	12	7	11.2217	0.71
12	11	13	1	11.1677	0.75
13	12	14	8	18.7373	0.67
14	13	15	10	14.4322	0.61
15	14	16	7	14.9935	0.72
16	15	17	9	13.3515	0.49
17	16	18	7	13.5216	0.6
18	17	20	9	14.4463	0.71

Figure 12: Train time and test scores

Conclusion

MODEL	ACCURACY
Decision Tree	65.8%
KNN Regression	73.4%
MLP Regressor	41.7%
Linear Regressor	60.2%
Random Forest	75.3%
Xgboost	71.1%

In conclusion Random Forest outperformed the other five models in performance and is used for Load prediction.

Ways to Improve Model Accuracy

- ➢ Remove irrelevant features
- ➢ Apply regularization to remove overfit
- ➢ Increase the amount of training data
- ➢ Ensemble learning: combining multiple models