Energy Load Prediction Using Multimodal Data from the CU-BEMS Dataset - Nayaab C (nic19)

Dataset Overview

The dataset contains data from several sensor types, including environmental sensors comprising of temperature, relative humidity percentage, and ambient illuminance. The dataset also contains records of electricity consumption from air conditioning units, lighting and plug loads in each of the zones in the seven-story office building.

Measurements are recorded at one-minute intervals for an eighteen-month period from July 1, 2018 to December 31, 2019.

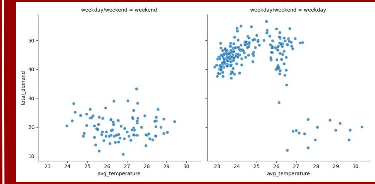
Problem Statement and Motivation

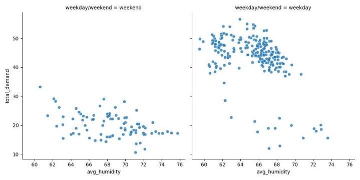
The objective of this project is to leverage the multimodal data provided by the dataset to attempt to predict energy loads through the application of time series forecasting techniques using multimodal learning models. In doing do, it would be possible to identify correlations between certain modalities, and perhaps form conclusions regarding the trend of energy consumption using domain knowledge.

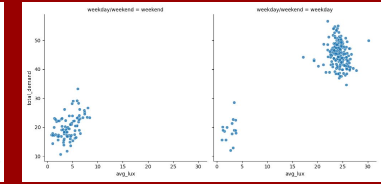
The primary motivation behind this project is to use the knowledge obtained in course to better understand how multimodal machine learning is used to analyze real world data trends. This project will also enhance machine learning software development skills.

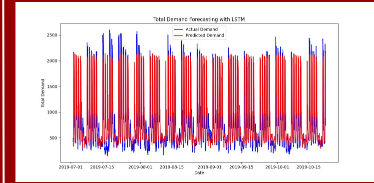
Methodologies and Results:

- Visualize data and create weekend, daily, hourly features - more detailed in report.
- Predict Energy Load using LSTM and Temporal CNN model architectures as well as Light Gradient Boosted Machine (GBM)
  - Early Fusion (red) - Train and Predict a single instance of the model using the concatenation of the data
  - Late Fusion (green) - Train and Predict multiple instances of the model for each modality, then take the weighted average of the output.

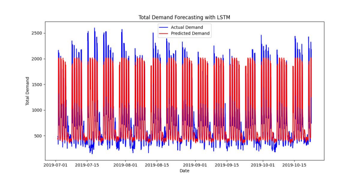




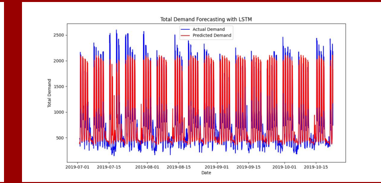




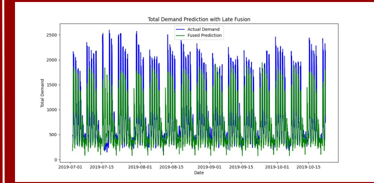
LSTM - Total Demand and Average Temperature



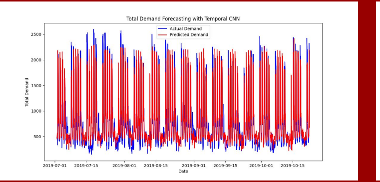
LSTM - Total Demand and Average Humidity



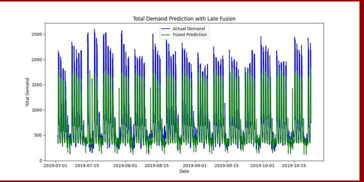
LSTM - Total Demand and Average Lux



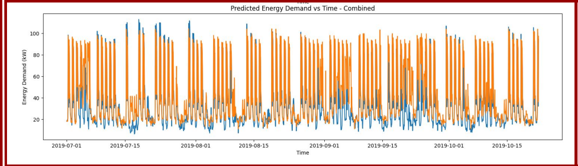
Temporal CNN - Total Demand and Average Temperature



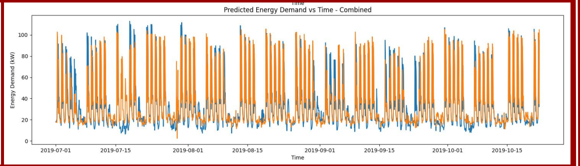
Temporal CNN- Total Demand and Average Humidity



Temporal CNN - Total Demand and Average Lux



Early Fusion - Total Demand and Average Temperature Light GBM



Early Fusion - Total Demand and Average Humidity - Light GBM

Actual Total Energy Demand (blue) vs Model Outputs

Metrics

LSTM metrics

	mape	nmbe	cvrsme	rscore
power	32.783744	1.526284	37.956275	0.739885
powerTemp	29.3004	-3.941245	30.159118	0.835458
powerHumidity	30.457842	-0.373308	36.980451	0.752609
powerLux	23.233798	-3.475086	24.87513	0.888064

Temporal CNN metrics

	mape	nmbe	cvrsme	rscore
power	27.718445	0.624057	29.190912	0.846152
powerTemp	22.960968	-1.11947	24.261864	0.893515
powerHumidity	24.615166	1.572131	26.378317	0.874127
powerLux	22.603626	-7.939803	21.740309	0.914499

Light GBM metrics

	mape	nmbe	cvrsme	rscore
power	26.97	0.85	29.09	0.85
powerTemp	39.17	-18.14	39.77	0.71
powerHumidity	23.82	1.01	28.7	0.85
powerLux	21.21	-8.92	19.59	0.93

Mean Absolute Percentage Error (MAPE)
How far, on average, the predictions are from the actual values

Normalized Mean Bias Error (NMBE)
Quantifies bias or systematic error in the predictions. Negative bias imply a more conservative estimate of values. Positive bias indicate a more generous estimation.

Coefficient of Variation of Root Mean Squared Error (CVRSME)
Provides relative measure of the error, while considering the scale of the observed values. The lower, the better the performance of the model

R-Squared Score (RSCORE)
Proportion of the variance in the target, which is explained by the predictions. Ranges from zero - meaning no variance - to one or perfect performance.

Energy load prediction could help serve the owners of the building by preparing ahead of time. This could be in the form of changing electricity sources to more efficient ones, preparing finances as necessary during specific timeframes (such as holidays and weekends), or through other methods available to them.