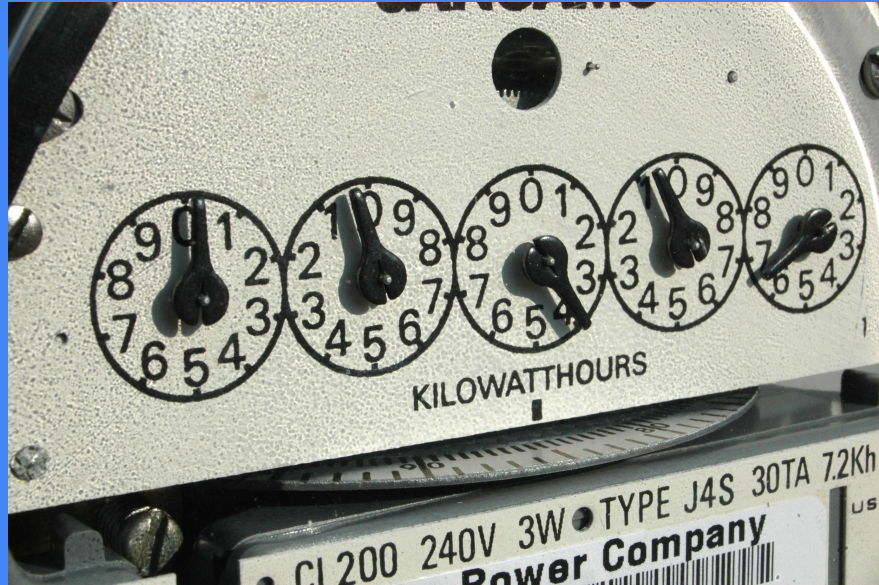


Can We Accurately Model Building Energy Usage?



By Justin Tyrrell

About me

- Outdoors
 - Kayaking
 - Hiking
 - Biking
- Concerts
- Environmental Studies
- Career
 - Energy Efficiency
 - Environmental Health
 - Data Science



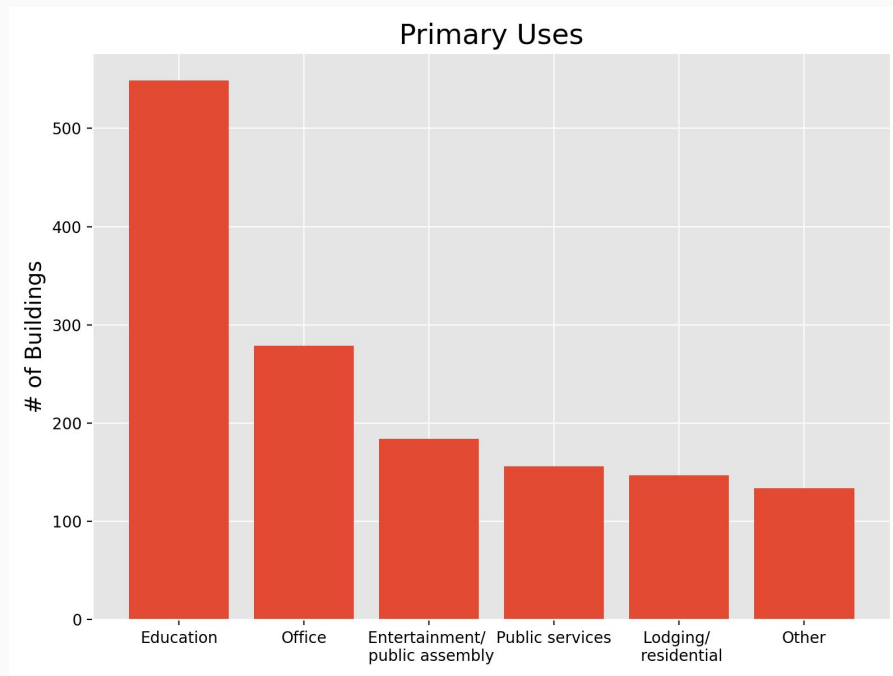
Why Should We Care?

- Current administration's focus on Climate Change
- Reducing energy consumption
- Energy efficiency
- Accurately modeling baseline usage

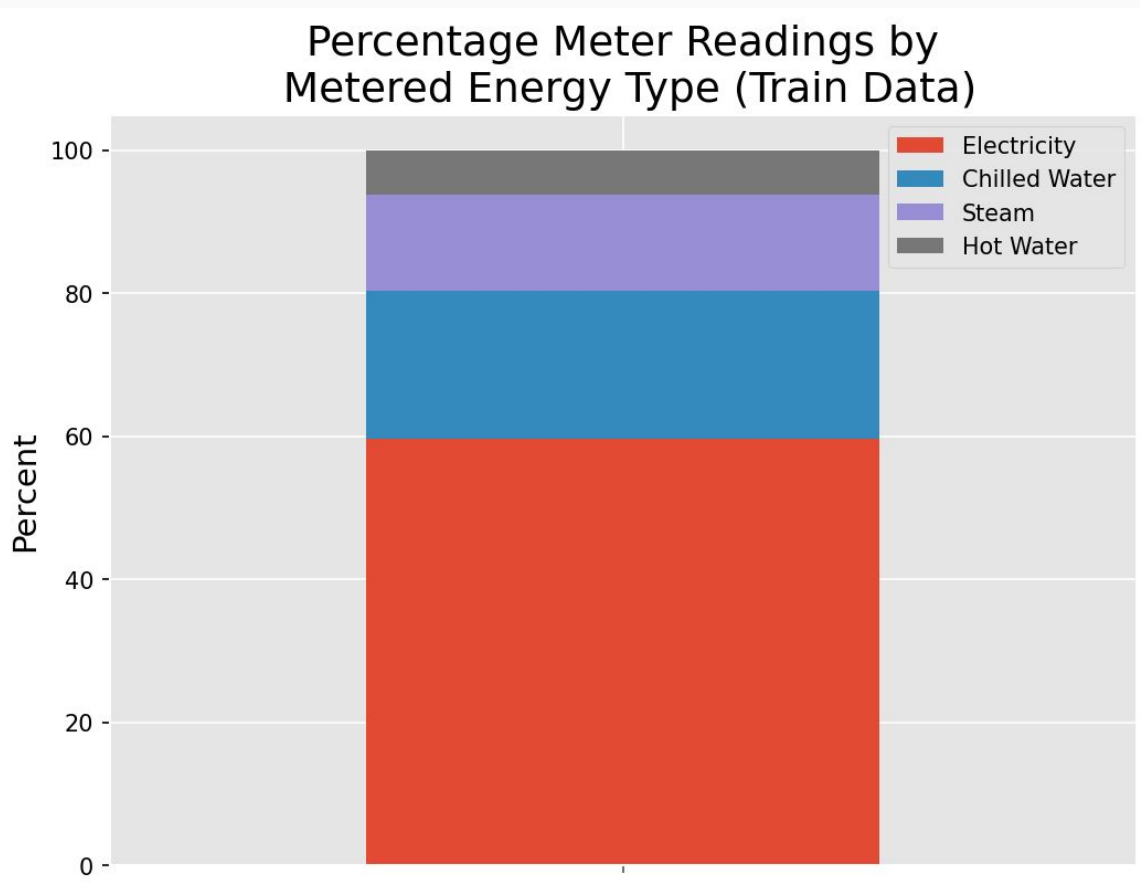


The Data

- Metadata
 - Includes unique building ID, site ID, primary use, and metered energy use type
- Energy usage and weather
 - Hour by hour for each building/site ID for 1 year
- Kaggle Competition
 - 20,000,000+ rows in train
 - 40,000,000+ rows in test



The Data



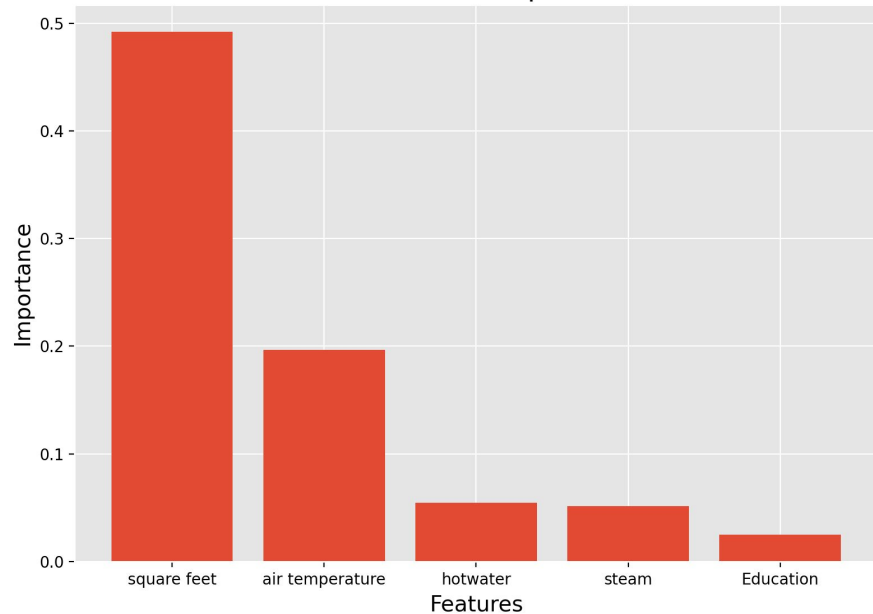
Building the Model

- Random Forest Regressor
 - Root Mean Squared Log Error (RMSLE)
- Split data into:
 - Water Usage Types
 - Electricity
- My RMSLE: 0.855
- Kaggle Winning RMSLE: 0.931

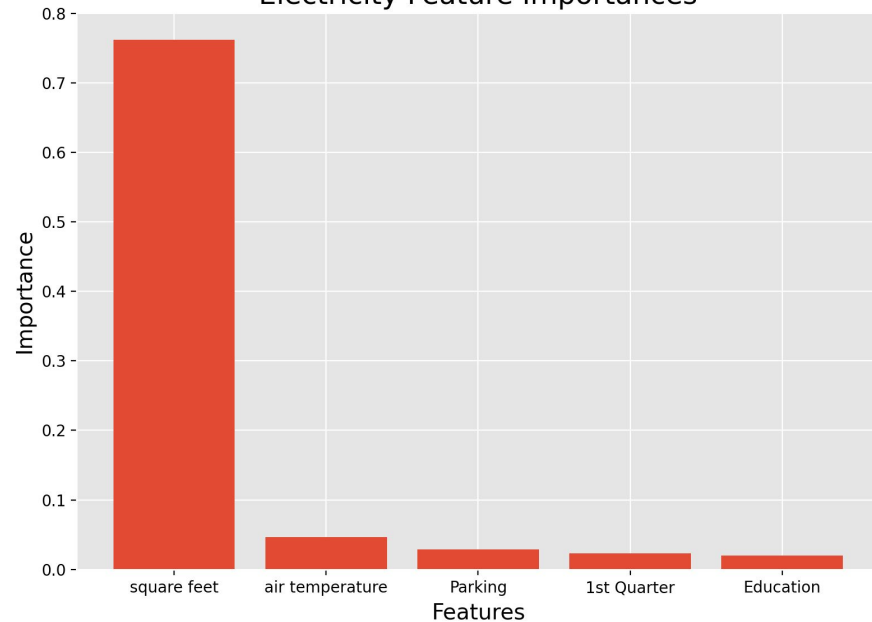


Feature Importances

Water Feature Importances



Electricity Feature Importances



Conclusions

- Preliminary results from modeling test data are promising
 - My RMSLE: 0.855
 - Kaggle Winning RMSLE: 0.931
- Square footage and temperature important to water metered energy types
- Square footage most important for electricity

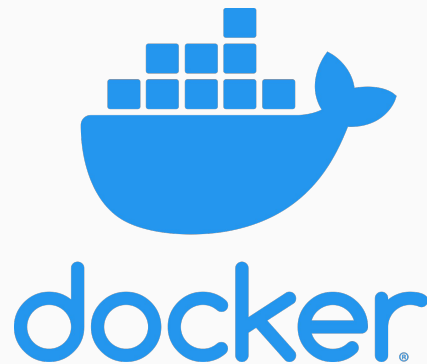
Questions?



Email: justin.d.tyrrell@gmail.com

LinkedIn: [in/justin-tyrrell/](https://www.linkedin.com/in/justin-tyrrell/)

Github: [/jtyrrell86](https://github.com/jtyrrell86)



Appendices

1. Sources
2. Future Direction
3. Data Visualizations
4. Additional Feature Importance Plots

Sources

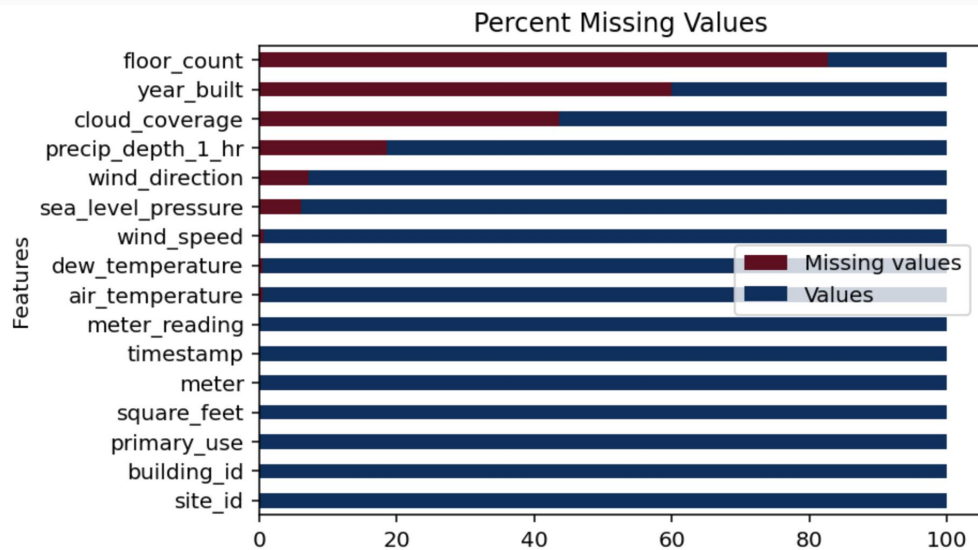
- Data - <https://www.kaggle.com/c/ashrae-energy-prediction/leaderboard>
- Energy stats - <https://www.eia.gov/>
- Biden admin's climate goals - <https://www.nytimes.com/live/2021/04/22/us/biden-earth-day-climate-summit>

Future Direction

- Tuning Hyperparameters
- Additional models
- Additional features
- Resolve AWS and docker issues
- Test dataset

Cleaning and Featurization

- Missing values
 - imputing missing temperatures
- Timestamp data, Primary usage, and meter columns
 - One hot encoding
- Usage conversion for site 0



Building Metadata

building_meta.csv		
Variable Name	Data Type	Description
site_id	int	Foreign key for the weather files.
building_id	int	Foreign key for training.csv
primary_use	object (string)	Indicator of the primary category of activities for the building based on EnergyStar property type definitions
square_feet	int	Gross floor area of the building
year_built	float	Year building was opened
floor_count	float	Number of floors of the building

Primary uses included Education, Entertainment/public assembly, Food sales and service, Healthcare, Lodging/residential, Manufacturing/industrial, Office, Other, Parking, Public services, Retail, Services, Technology/science, Utility, and Warehouse/storage.

Energy Usage and Usage Type

train.csv		
Variable Name	Data Type	Description
building_id	int	Foreign key for the building metadata.
meter	int	The meter id code. Read as {0: electricity, 1: chilledwater, 2: steam, 3: hotwater}. Not every building has all meter types.
timestamp	object (string)	When the measurement was taken
meter_reading	float	The target variable. Energy consumption in kWh (or equivalent). Note that this is real data with measurement error, which we expect will impose a baseline level of modeling error. UPDATE: as discussed here, the site 0 electric meter readings are in kBTU.

Over 20 million rows of data

Weather Data

weather_[train/test].csv		
Variable Name	Data Type	Description
site_id	int	
timestamp	object (string)	When the measurement was taken
air_temperature	float	Degrees Celsius
cloud_coverage	float	Portion of the sky covered in clouds, in oktas
dew_temperature	float	Degrees Celsius
precip_depth_1_hr	float	Millimeters
sea_level_pressure	float	Millibar/hectopascals
wind_direction	float	Compass direction (0-360)
wind_speed	float	Meters per second

Cleaned Data

	square_feet	meter_reading	air_temperature	q1	q2	q3	q4	afternoon	early_morning	evening	...	Office	Other	Parking	Public services	Retail	Services
0	7432	0.0	25.0	1	0	0	0	0	1	0	...	0	0	0	0	0	0
1	2720	0.0	25.0	1	0	0	0	0	1	0	...	0	0	0	0	0	0
2	5376	0.0	25.0	1	0	0	0	0	1	0	...	0	0	0	0	0	0
3	23685	0.0	25.0	1	0	0	0	0	1	0	...	0	0	0	0	0	0
4	116607	0.0	25.0	1	0	0	0	0	1	0	...	0	0	0	0	0	0