Predicting Volatility in US Electricity Prices

Miles Franklin, David Harper,
Adam Goldstein, Himanshu Ghritalhre

Importance

Electricity price volatility is important to understand

- As the U.S. starts looking more into renewable energy, it is important to understand what impacts it could have
- Understanding this volatility could help with a smooth transition into renewable energy sources
- This understanding could also help with decision making for non-renewable energy sources
- Decide on which parties to incentivize new adoption in where price volatility won’t be as effected. I.e Residential, Commercial, or Industrial customers.
- Which states could feel the least impact from changes

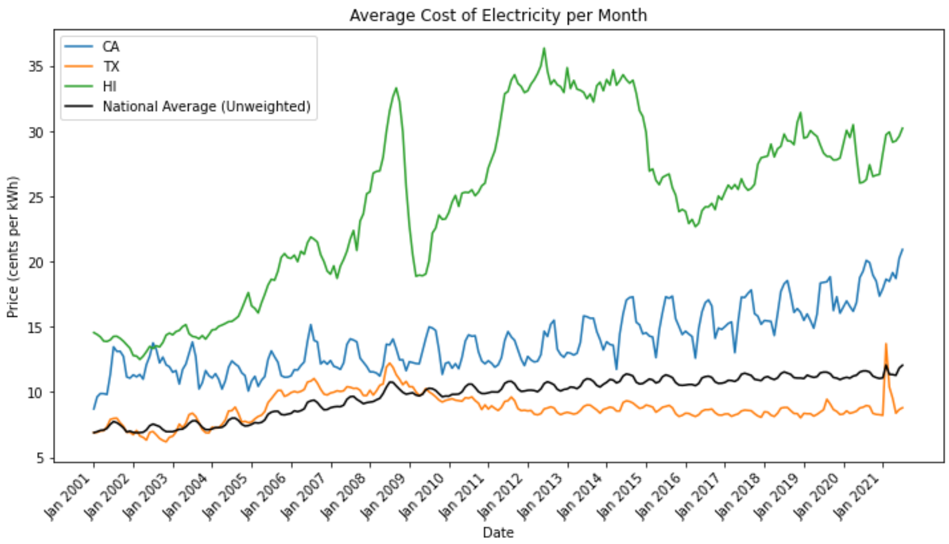
UMBC

Overview

Target

The goal of this project is to produce a model predicting the price volatility of electricity for a given **year**. We measure the volatility using the Coefficient of Variation, to normalize the data and facilitate state to state comparisons.

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Mean}}$$



UMBC

Data Sources (2001 - 2021)

Power Plants, Electricity Sales, and Electricity Accounts

eia

Independent Statistics & Analysis
U.S. Energy Information Administration

US State Population Estimates

United States Census Bureau

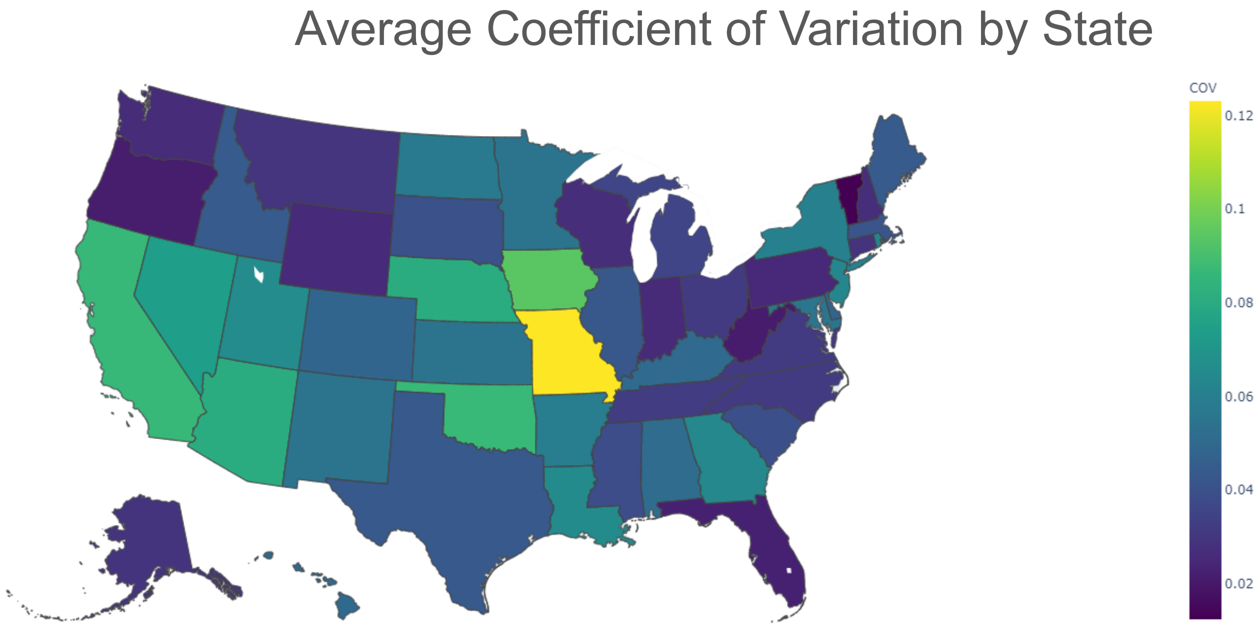
Future Contracts

yahoo! finance

Weather and Drought

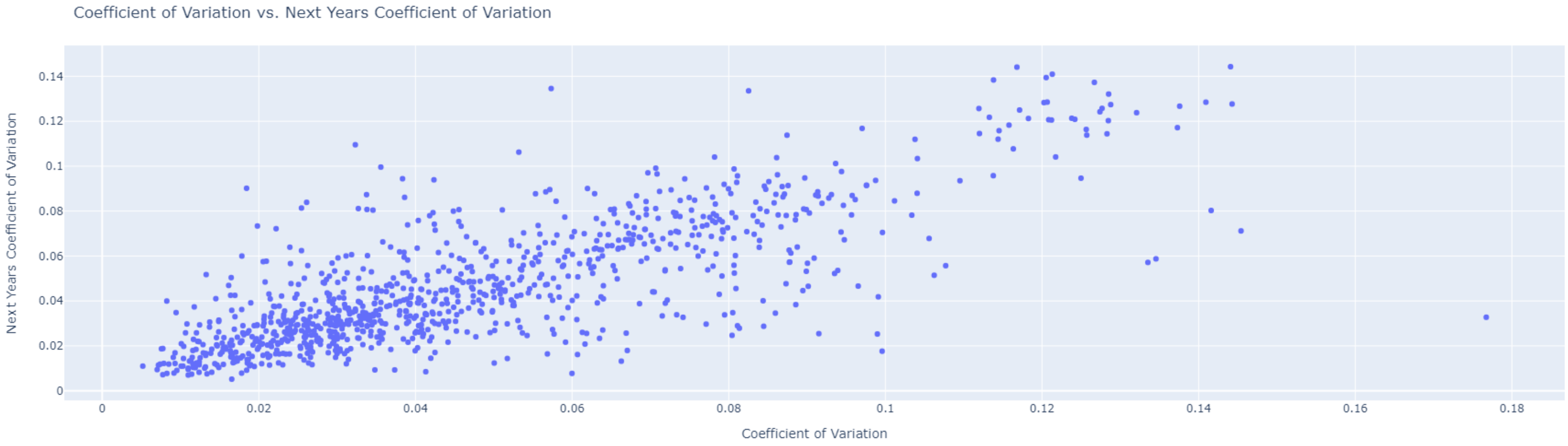
Google Earth Engine

Exploratory Analysis



Exploratory Analysis

- The most highly correlated variable with the coefficient of variation is the prior years coefficient of variation



UMBC

Modeling Pipeline

Feature Engineering

Used aggregations and calculations to create new features with different information

Scaled numerical columns to fit with in [0,1]

Model Selection

All of our potential algorithms are organized cleanly into one place for quick access

Model Training

Log Results

neptune.ai

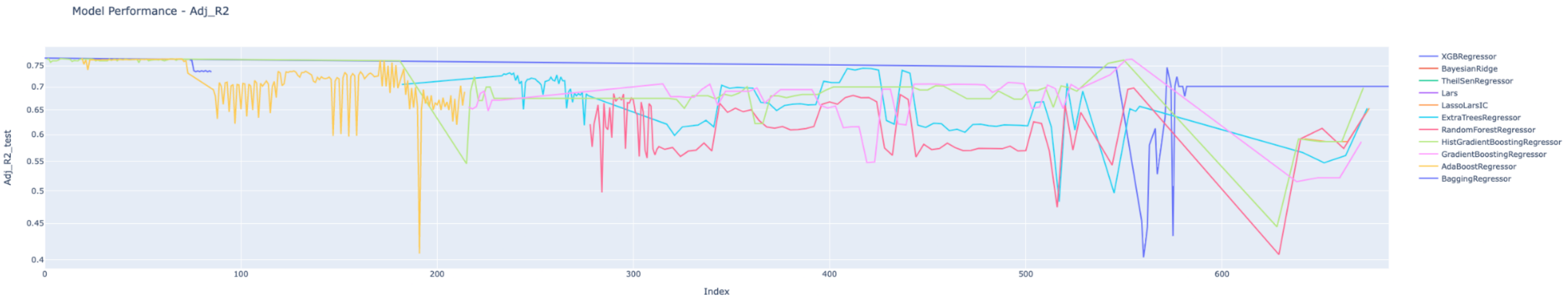
Compare Models

We compared Models using Adjusted R^2, a metric that shows the percent of variance explained by the model. While accounting for model complexity.

Neptune AI

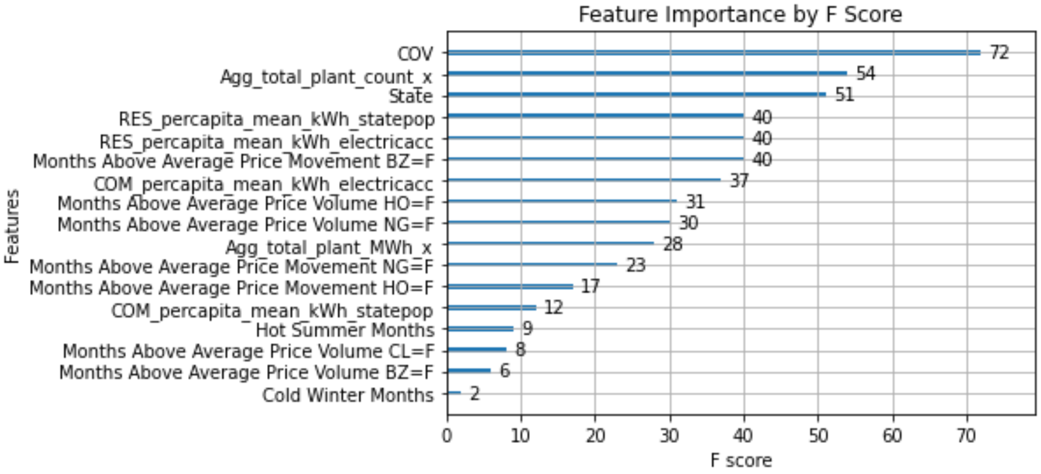
Unique Experiments - **217**
Feature Pipelines - **15**

- 17 unique models
- 7 Ensemble
- 10 Linear



Final Model Choice

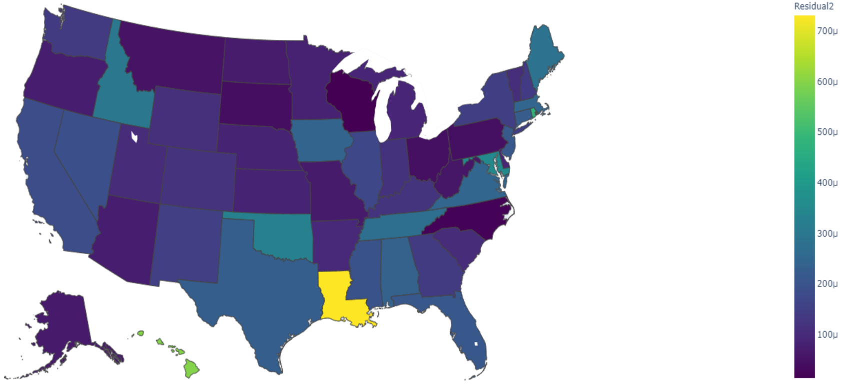
- XGBoost Tree Model
- Was able to account for the most variance in our data
- (Adjusted R^2=0.77)
- Found that the most important features were:
  - The variance of the previous year
  - The total plant counts
  - The State of interest
  - Mostly metrics reflecting supply and demand of electricity



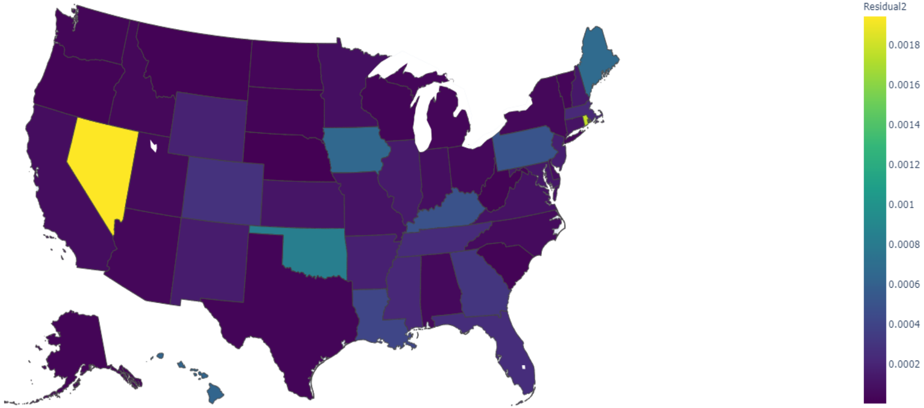
Cust_Model	base_estimator	Adj_R2_test
XGBRegressor	nan	0.768890
HistGradientBoostingRegressor	nan	0.766902
BaggingRegressor	HistGradientBoostingRegressor	0.763546
AdaBoostRegressor	LassoLarsIC	0.763318
ExtraTreesRegressor	nan	0.713314
GradientBoostingRegressor	nan	0.695125
RandomForestRegressor	nan	0.685285

Model Performance by State

Training MSE by State



Test MSE by State



Future Directions

Different Data Sources

- EIA was a helpful starting point, but was limited by the number of recorded years
- The model would likely benefit the most from having more data to learn from

Feature Reduction

- Many of the final models use all ~120 features
- Feature reduction algorithms would help reduce the complexity of the models, and improve the performance without the need for new data

Transforming into a Classification Model

- We began to explore classification models
- Potential buckets for target variable
    - Basic increase or decrease in COV from previous year
    - Buckets of Low, Medium and High COV

Thank You!