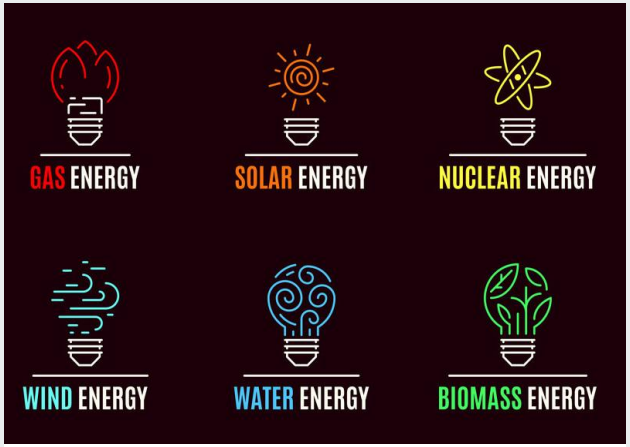
Prediction of Energy Consumption in California



TEAM 1

Frank Dong, Yang Weng, Nima Hendi, Sean Lee

University of California, Irvine collaboration with Accenture pls

Summary

Industry Energy Consumption Predictor

(1) collect historical data on energy consumption and related potential data
(2) preprocess and clean the data
(3) identify relevant features and engineer new ones
(4) develop and train machine learning models, eg MLP
(5) evaluate the performance of the models and fine-tune them as necessary.

Ultimately, the goal is to provide accurate predictions that can help make informed decisions and optimize energy usage.

Technical Approach

- Combine the industry energy related data sets.
- Industry energy consumption data EDA (Exploratory data analysis).
- Logistic and linear model, MLP neural network models comparison.
- Plot prediction with real data.

Data Sets

url for the datasets: https://www.eia.gov/totalenergy/data/monthly/

U.S energy summary

U.S energy consumption summary

Industry energy consumption by category

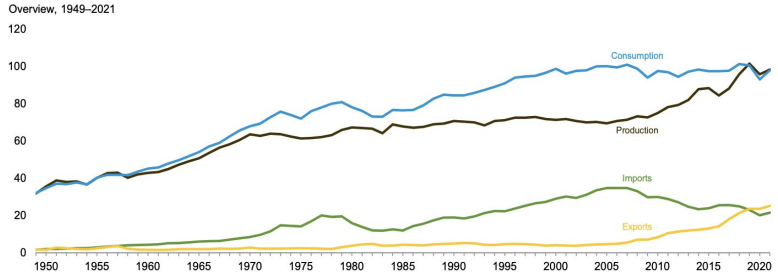
Energy trade by category

Energy prices by category

Industry emissions by source

	MSN	YYYYMM	Value	Column_Order	Description	Unit
0	FFPRBUS	194913	28.740479	1	Total Fossil Fuels Production	Quadrillion Btu
1	FFPRBUS	195013	32.553205	1	Total Fossil Fuels Production	Quadrillion Btu
2	FFPRBUS	195113	35.782118	1	Total Fossil Fuels Production	Quadrillion Btu
3	FFPRBUS	195213	34.964662	1	Total Fossil Fuels Production	Quadrillion Btu
4	FFPRBUS	195313	35.338361	1	Total Fossil Fuels Production	Quadrillion Btu

Data frame of US Energy Summary



Graph provided on the website

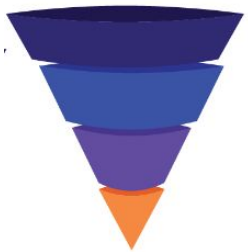
Data Sets (Cont.)

MSN	YYYYMM	Value	Column_C	Descriptic	Unit
NGWPUUS	194913	0.06	1	Natural Ga	Dollars per Thousand Cubic Feet
NGWPUUS	195013	0.07	1	Natural Ga	Dollars per Thousand Cubic Feet
NGWPUUS	195113	0.07	1	Natural Ga	Dollars per Thousand Cubic Feet
NGWPUUS	195213	0.08	1	Natural Ga	Dollars per Thousand Cubic Feet
NGWPUUS	195313	0.09	1	Natural Ga	Dollars per Thousand Cubic Feet
NGWPUUS	195413	0.1	1	Natural Ga	Dollars per Thousand Cubic Feet
NGWPUUS	195513	0.1	1	Natural Ga	Dollars per Thousand Cubic Feet

Data frame of Energy prices by category

MSN	YYYYMM	Value	Column_C	Descriptic	Unit
CLICEUS	197301	33.48	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197302	30.834	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197303	31.625	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197304	31.119	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197305	31.643	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197306	30.006	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197307	29.437	1	Coal Indus	Million Metric Tons of Carbon Dioxide
CLICEUS	197308	29.03	1	Coal Indus	Million Metric Tons of Carbon Dioxide

Data frame of Industry emissions by source



Linked together through the date attribute, “YYYYMM”

Experiments and Evaluation

- Use MSE to evaluate the model performance.
- Use cross validation to avoid overfitting and underfitting. Divide training and testing in time series.
- Modified training data along training to build better models, since our data cover a long period and old data might be noise for precise prediction.

Experiments and Evaluation (part 2)

Provide a brief and clear description of how you plan to evaluate the results of your project. For example, if you are doing classification, you should consider metrics such as classification accuracy and precision-recall, in addition to having access to training data. Will you use cross-validation, or does your data set(s) come with a fixed train-test partition? For unsupervised learning tasks like clustering or topic modeling, you may have to do some research to see how evaluation is done on these tasks. For some projects you may even have to do some user studies for evaluation, e.g., present users with results from Algorithm A and Algorithm B, using the same input data for each algorithm, without telling the user which algorithm is which, and have them select the one they prefer. Or your evaluation may be more qualitative in that you hope to generate insights about a particular problem.

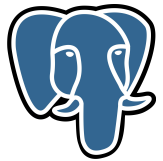
Software and Programming Libraries

Software

- Visual Studio Code \ JetBrains
- PostgreSQL Desktop+ Shell
- GitHub Desktop
- Anaconda Navigator
- Jupyter Notebook

Programming Libraries

- Python
  - Pandas
  - Numpy
  - Matplotlib
  - PyTorch
  - Sklearn
- R Studio
  - Tidyverse
  - Tidypredict
  - dplyr




Code we will write: appropriate schemas, pipelines, web application, prediction implementation

Milestones

- Winter
  - Weeks 8-10 Data sets Collection and Combination
- Spring
  - Weeks 1-2 Data sets EDA
  - Weeks 3-4 Statistic Model fit, and Machine learning Model fit
  - Weeks 5-6 Model validation and visualization
  - Weeks 7-8 Model analyze
  - Weeks 9-10 Final report