ELA - Energy by Location in America

Sai Varun Aduru, Heidi Nelson, Khushmeen Sakloth, Ethan (Fangqi) Wang
github.com/DIRECT-Energy-Storage/ELA

Objectives

- To develop an educational toolkit for studying the distribution of energy in the United States.
- To provide new visualizations of energy generation and storage facilities.
- To explore the geographic distribution of energy generation and storage facilities by technology type, using machine learning models.

Data Sources

Global Energy Storage Database (U.S. Department of Energy)

http://www.energystorageexchange.org

Information about 678 U.S. energy storage facilities, including:

- Facility name and location
- Type of energy technology (electro-mechanical, pumped hydro storage, electro-chemical, or thermal storage)
- Power in kW and energy storage duration
- Primary use case

Emissions & Generation Resource Integrated Database (U.S. Environmental Protection Agency)

https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid

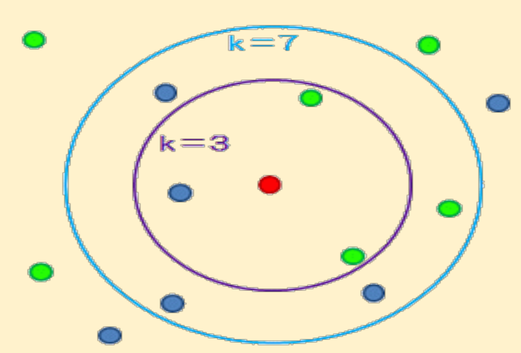
Information about 8462 U.S. energy generation facilities, including:

- Facility name and location
- Type of energy technology (gas, oil, coal, biomass, nuclear, geothermal, solar, wind, hydro, or other)
- Rated capacity in MW and actual production in GWh

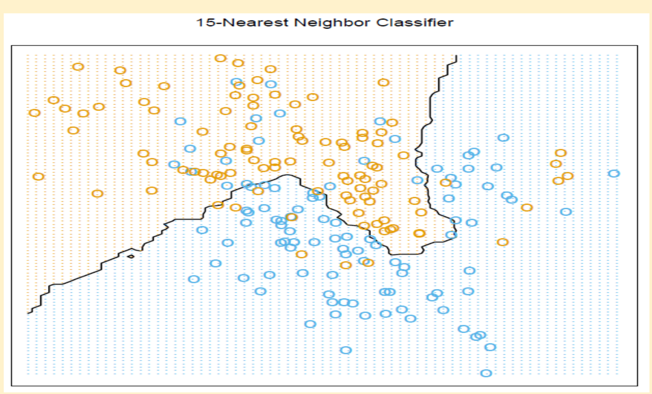
Data Science Method: K Nearest Neighbors

- K nearest neighbors, or KNN, is a simple and intuitive machine learning algorithm. It is a non-parametric method (parameters are determined by the training data, not the model). For classification, an object is assigned to the class most common among its *k* nearest neighbors.

- KNN offers several adjustable inputs, most obviously the value of k. Using too small of a value of k can lead to a model that is very specific, but not generally meaningful. On the other hand, using too large of a value of k can lead to an overly general and less accurate model. Beyond the choice of k, various distance metrics can be used to determine which neighbors are nearest. Also, after the nearest neighbors are selected, their influence on the assigned class can be weighted equally or based on their distance to the point under consideration.



Schematic of KNN Algorithm
Image Source: http://www.news2u.net/releases/34629

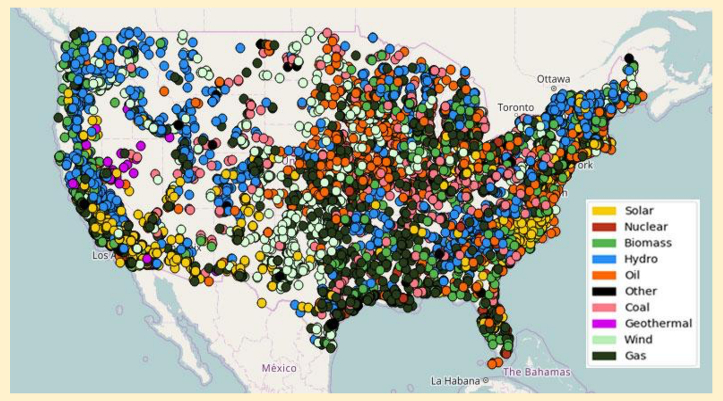


The elements of statistical learning: data mining, inference, and prediction, figure 2.2

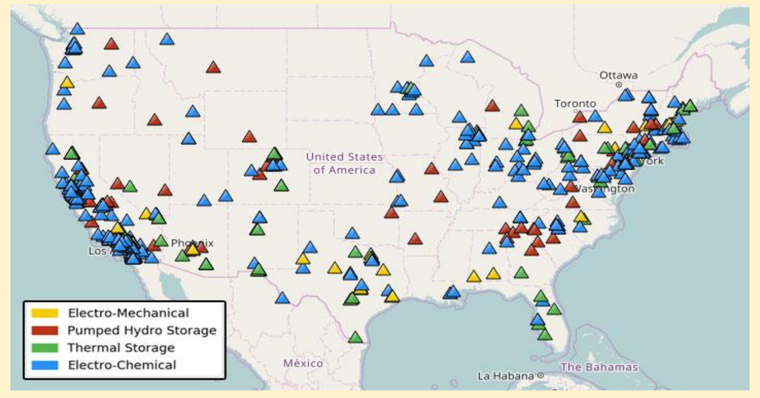
- In our model, nearest neighbors are determined by actual physical distance. The model takes in latitude and longitude coordinates, finds the closest facilities in the data set, and predicts the energy type for the input location based on the energy types of the closest facilities. The overall accuracy of this model can show whether geographic proximity is a good predictor of energy generation or storage type. It also provides an interesting visualization of dominant energy types in different regions.

KNN Exploration & Visualization

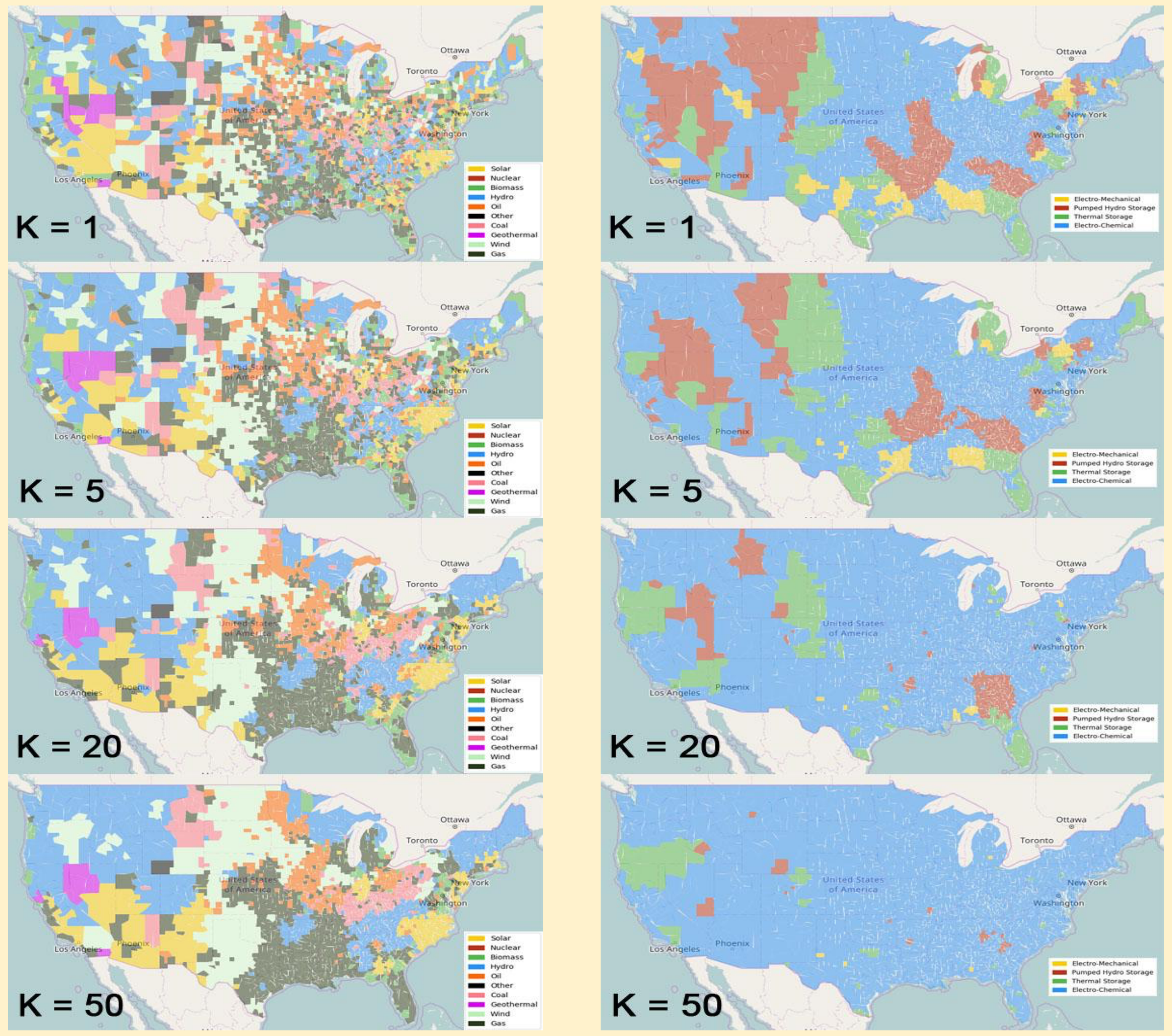
Energy Ggeneration Facilities



Energy Storage Facilities



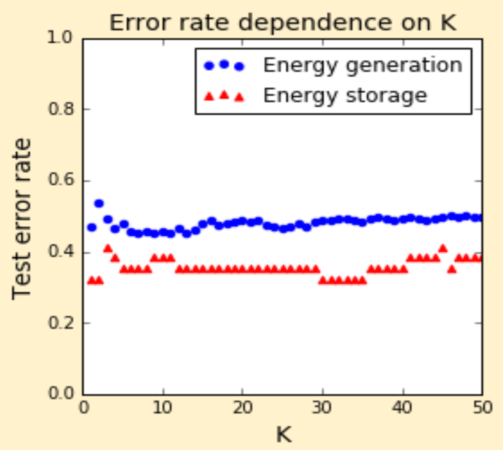
- Using a K nearest neighbors model based on these energy facilities, we predicted the primary energy generation and storage type for each U.S. county:



For K = 1, the model is simply finding the type of the closest energy facility for each location. With increasing K, some apparent geographic trends emerge. For example:

- Hydroelectric power is common in the Pacific Northwest
- Many wind farms are located in the Midwest and Great Plains regions
- The Rust Belt area (Great Lakes/Midwest) contains many coal plants
- Electrochemical energy storage is most common nearly everywhere in the U.S.

The accuracy of a KNN model can be evaluated by dividing the data into training and testing sets and using the training data to predict the classification of the points in the testing data set. The error rate (frequency of incorrect classifications) can quantify the performance of the model. For a KNN model trained on 95% of each data set, the error rate in predicting the energy type for the other 5% of facilities is ~50% for energy generation and slightly less for energy storage. KNN does not appear to be a very accurate model for predicting energy type.



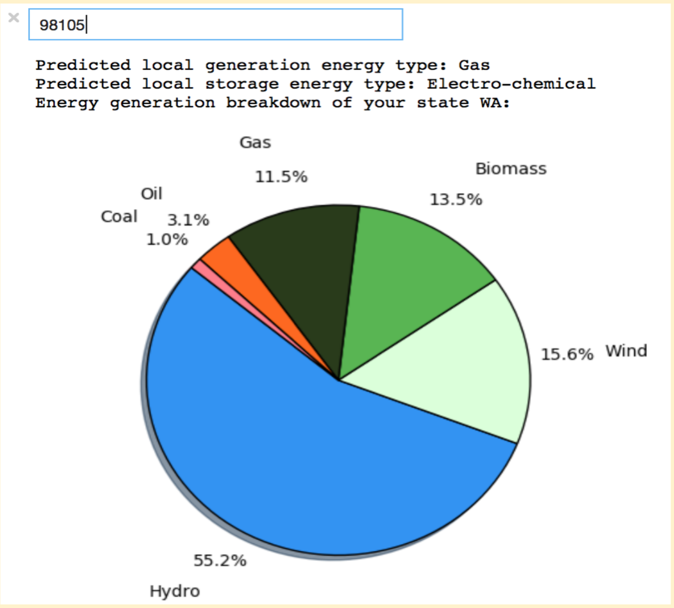
Further exploration of the KNN model is available in the ELA python package.

User Interface

ELA has two main components: zip code lookup and map visualization.

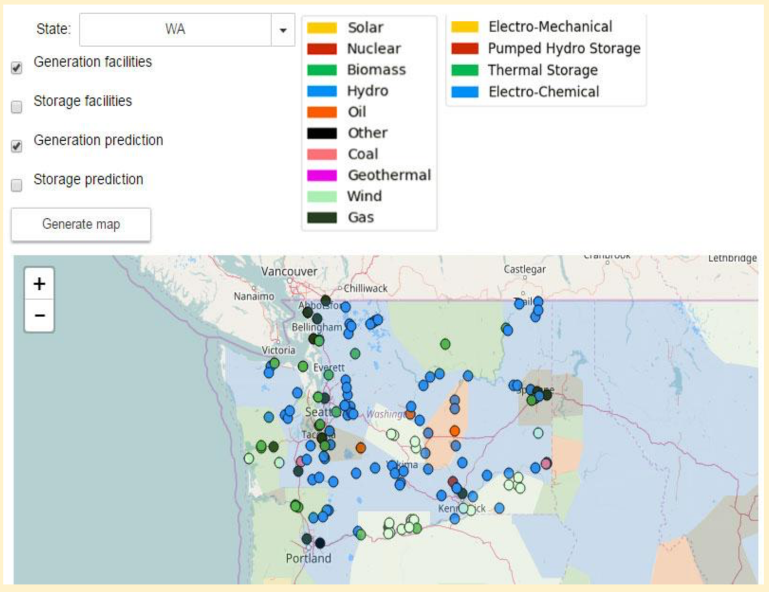
Zip Code Lookup

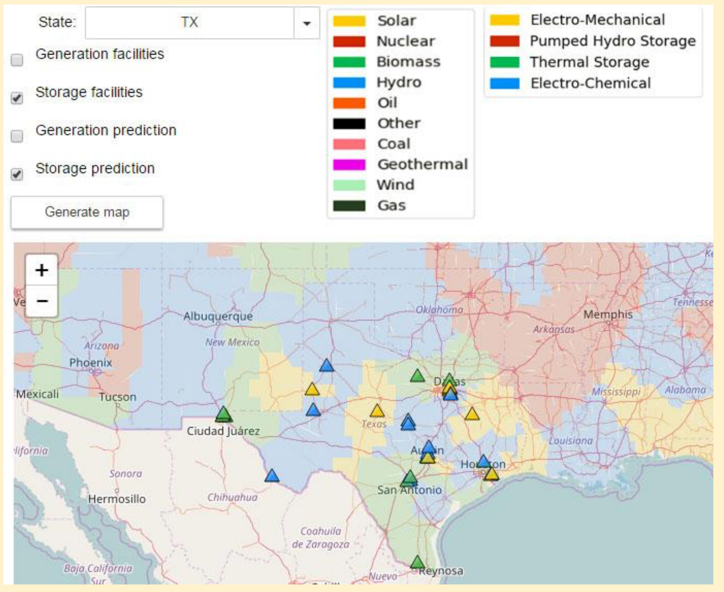
The user can enter any zip code to learn about the closest energy generation and storage facilities, predicted energy types based on nearby facilities, and comparison with the distribution of energy facilities in their state and in the U.S.



Map Visualization

The user can visualize energy generation or storage facilities in any state, or the predicted energy generation or storage types across the entire U.S. (based on the KNN model).





Challenges and Limitations

- This approach does not take into account how electrical grids and utilities actually work on a regional level, instead focusing on the closest facilities.
- A major issue with our model is that it only considers the number of facilities, not the size of facilities, when predicting energy types. The capacity or production of the energy facilities is not considered at all, which can lead to many inaccurate conclusions about energy usage.
- In the storage data, electrochemical facilities are a majority, which limits the usefulness of KNN. If we predicted that all new energy storage facilities would use electrochemical technology, we would be more accurate than the KNN model.
- Overall, the KNN model does not appear to capture most of the factors influencing the types of technology used in energy facilities. Geographic proximity to existing facilities is not a very good proxy for these other factors (resource potential, economic considerations, etc.).

Acknowledgements

- David Beck and Jim Pfaendtner for the opportunity to participate in DIRECT.
- The e-Science Institute
- UW Clean Energy Institute