

Capstone Project Proposal Template

World Energy (Renewable & Non-renewable) Consumption

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?

The business problem being solved is trying to identify which energy sectors are more likely to increase in demand thus resulting in being more valuable assets to investors. Also, it allows global market competitors to focus on making products more suitable to these energy sectors.

- What industry/realm/domain does this apply to?

This project applies to the energy sector/industry.

- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)

The motivation for this project is to understand the rate at which energy is being consumed and which energies are favorable and non-favorable to the largest populations in the world.

Data Understanding

- What data will you collect?

The data collected is from a major Kaggle project that has collected world energy consumption measurement details in the same format.

- Is there a plan for how to get the data (API request, direct download, etc.)?

Fortunately, the data is immediately downloadable/accessible in csv format after downloading and unzipping the folder from Kaggle.

- What are the features you'll be using in your model?

The features that will mainly used in the model will be years (21st century), countries (China, India, United States), wind consumption, gas consumption, nuclear

consumption, solar consumption, oil consumption, coal consumption, & hydro consumption.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?

The data will need to be prepped for removing null values and for any inconsistent data types that need to be formatted properly.

- What are some of the cleaning/pre-processing challenges for this data?

Some of the biggest challenges for this data was that some of the major players in the energy sectors (Russia, Nigeria) do not have public data, thus making it very hard to find the actual world data for certain major energy countries.

Modeling

- What modeling techniques are most appropriate for your problem?

Regression techniques are most appropriate for this problem. For model evaluation, techniques like Cross Validation (stratification) would seem to fit best.

- What is your target variable? (Remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)

The target variable will be future wind consumption rates measured in Terawatt hours (TWh).

- Is this a regression or classification problem?

This will mostly be a regression problem as it aims to project usage amounts. Although, in a sense it could also be perceived as a classification problem because it allows investors to shortlist which energy sectors they should invest in for the future.

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?

The main metric that will be used to determine success is model accuracy score.

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?

The modeling algorithms that I am planning to use are linear regression, decision tree regressor.

Sources

[Energy Consumption Prediction with Machine Learning | Aman Kharwal \(thecleverprogrammer.com\)](#)

[World Energy Consumption | Kaggle](#)