# ML Project Proposal: OASIS

Optimal Allocation for Sustainable Investment Strategies

Prayash Joshi, Charles Bruner, Arsh Mohamed Siddiqui
Virginia Tech

Proposed Date: October 3, 2023

## Executive Summary

The United States has set forth various sustainability goals with a target set for the year 2050. These goals, although admirable, have been critiqued for their general nature (including questionable attainability and nationwide generalizations), and many critics argue that they may not be fully achievable by the stipulated deadline. To address this challenge, our project aims to design a machine learning model that can generate optimal sustainable investment plans specific to each state in the U.S. This alternative recognizes that each state possesses different energy capacities and capabilities, thus our model will tailor its recommendations accordingly, providing a holistic road map to sustainability by 2050.

## 1 Preliminary Plan

Our approach involves studying the unique energy landscape of each state, comprehending the current energy consumption, available renewable resources, and understanding the potential of each sustainable practice. A machine learning model will be built from scratch to evaluate these parameters and make informed decisions on the most optimal energy mix. The overarching goal is to help guide states toward making the most effective sustainable practice investments for a given period. The next step moving forward is to determine a concrete metric to improve, measure, analyze, and test. This may involve developing a novel scoring procedure for the appropriate sustainable investment combinations. Then we will start to review the types of data available to us, like historic time series data and geospatial data that reflect the energy capabilities of individual states. We would then intend to combine our data, prepossess it, and perform feature engineering. Note, this is what we hope to accomplish across the next month. The plans beyond that are listed below in our milestones.

## 2 Milestones

- **Data Collection and Preprocessing:** Gather data on energy capacities, consumption, and available resources for each state.
- **Feature Engineering:** Extract and create relevant features from the data-set. Features might include state population, state GDP, historical energy consumption, available renewable resources, etc.
- **Model Architecture Design:** Design the architecture of the model, considering factors like model type, layers, and neurons (in case of neural networks), or the depth (in case of decision trees).
- **Training and Validation:** Split the dataset into training, validation, and test sets. Train the model on the training set and validate its performance.
- **Optimization:** Refine the model by tuning hyperparameters, regularization techniques, and other relevant strategies to improve performance.
- **Evaluation:** Test the model using custom A/B Testing and Holdout Set Evaluation.

- **Deployment:** Integrate the model into a user-friendly interface, create beautiful visualizations and make it public to all users on the internet.

## 3 Roles

- **Lead ML Engineer/Researcher**: Prayash Joshi
- **Lead Data Scientist/Researcher**: Arsh Siddiqui
- **Lead Data Engineer/Developer**: Charles Bruner

## 4 Paper List

- P. Asha, et al., Role of machine learning in attaining environmental sustainability, Energy Reports, Volume 8, Supplement 8, 2022, Pages 863-871, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.09.206.

  This presents a specific determination of attainability of sustainable development goals in relation to solar panels. Provides an idea of what a model involving determining sustainability in particular relation to development goals might look like.

- Mehrbakhsh Nilashi, et al., Measuring sustainability through ecological sustainability and human sustainability: A machine learning approach, Journal of Cleaner Production, Volume 240, 2019, 118162, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2019.118162.

  This attempts to assess the performance of certain countries sustainability goals. This can be applied to specific states to see where they are succeeding and where they are lagging behind, which would provide insight into potential sustainable targets per state.

- Tanveer Ahmad, et al., Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment, Energy, Volume 158, 2018, Pages 17-32, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2018.05.169.

  This paper measured and predicted the energy usage of buildings in New England one year ahead. Understanding the energy consumption of buildings is a part of the U.S.'s 2050 sustainability goals, thus understanding how certain states can relate or improve on this metric would be desired.

- You Han, et al., Estimation of Corporate Greenhouse Gas Emissions via Machine Learning, 2021, 2109.04318, arXiv, https://arxiv.org/abs/2109.04318

  The amount of greenhouse gas emissions done by corporations is an important metric to consider, especially with respect to the diversity of companies across the fifty states. Understanding how best to estimate their output would thus be useful for determining the state-wide output of greenhouse gases. This would also be critical for classification of greenhouse gases by industry.

- Ümit Ağbulut, Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms, Sustainable Production and Consumption, Volume 29, 2022, Pages 141-157, ISSN 2352-5509, https://doi.org/10.1016/j.spc.2021.10.001.

  This covers transportation related emissions in Turkey, and forecasts its growth and change over time. Transportation related emissions is another one of the U.S.'s 2050 sustainability goals, thus understanding the sorts of models we could use to predict emissions of transportation is additionally useful.