**CIND 820:**

**Project Abstract**

By Aidan Kowolik

Student No. 500700734

Course Supervisors: Ashok Bhowmick, Ceni Babaoglu, Sedef Akinli Kocak, Tamer Abdou, Uzair Ahmad, Zekiye Erdem

Due Sept, 25, 2023

Course Code CIND 820 D1H

**Themes:**

The primary theme of the project will be using predictive analytics to predict future emissions and to compare countries against each other. Through this analysis, we hope to see what are the factors causing emissions and how these countries can lower them. By knowing these factors, recommendation systems could be put into place to advise countries on how they could lower their emissions.

**Dataset:**

The database that I’ll be using is from Kaggle and the title of it is; Global Data on Sustainable Energy (2000-2020).[[1]](#footnote-0) This database was obtained from the website with the Attribution 4.0 International (CC BY 4.0) license. The historic data is obtained from countries for the last twenty years with 176 unique values with twenty-one attributes for each row of data.

**Methods:**

**Techniques:**

The primary programming language that will be used is Python and the techniques that will be used are Linear regression, K-means and Decision Tree analysis.

Linear regression will be used to predict future emissions produced based upon the factors measured with the module Sklearn with the function linear\_model.LinearRegression.

K-means is used to cluster countries together to see what contributing factors toward their emissions are and the areas that need improvement or require help to lower their emissions.The module used for K-means will also be from Sklearn with the k-means function.

Decision tree will be able to classify the most important attributes from the dataset that will be a predictor for emissions. The module that will achieve this will be the decision tree classification from Sklearn as well.

**Research Questions:**

The major research questions that I will answer in this project is to predict the various countries' emission levels from the historical data and compare it against their goals. The project will identify which countries need help or are on track and to see what are the contributing factors for both scenarios.

By finding out what factors affect their emissions we can find what each country needs in terms of policy evaluation and planning to make progress towards their emissions goals.

During this project we will be able to identify countries that are in need of energy infrastructure using clustering to understand how sustainable energy could play a role in their development.

1. Link to the database: https://www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy [↑](#footnote-ref-0)