Research Question: Global warming and climate change have become daily issue topics, all because of carbon emission. Among many factors, one of the issues that is causing an increase in carbon emission is the energy consumption per household. *Is it possible to reduce carbon emissions with the usage of an energy consumption prediction model to predict the amount of renewable energy from wind turbines that is needed to lower carbon emission levels in the Netherlands?* This research question addresses the issue of increasing carbon emissions and aims to solve it by analyzing various data sets from factors, such as the amount of energy generation from turbines on a monthly basis, in comparison to the demand of energy, that will help us understand how these factors are affected and come up with a suitable solution to decrease carbon emission levels. The goal is to produce a model that is capable of predicting the amount of energy consumption per household on a monthly basis, in order to run and produce the most efficient amount of renewable energy.

Business Understanding: The social cost[9] of carbon pollution mentioned above does not include the cost of scientific and economical impacts that are also produced from each household. Although, it is possible to lower the amount of carbon pollution by providing more sources of renewable energy, such as electricity through wind power. The average household in the Netherlands 2016 consumed 2,910 (kWh) of electricity [2]. Of this amount, the energy consumption of renewable energy in comparison to all the sources of energy that was consumed in the average household was only 4.4%, whereas natural gas and electricity were at 75% and 15% respectively[2]. Therefore, by reducing the social loss caused due to carbon emission by processing coal and natural gases, among other sources of energy, this can help reduce the social cost of carbon emission for the customers and the companies supplying energy.

The model will require historical data of the amount of carbon emission produced in relation to monthly consumer energy consumption trends. The required data will be collected from power plant databases and open data from online sources, for example the Rijksoverheid [8]. However, there is a possibility that the data in need for collection may be restricted to access, due to the constantly changing law of sharing data. In case of this mishap, data sets may need to be bought under a license

In order to interpret the knowledge retrieved through the data properly, it would help to have experts that study carbon emission and renewable energy related professionals would be other invaluable sources of information that would help answer the research question. Moreover, data scientists would be needed to help analyze the data and process it to make it understandable. In addition, to set up the model, machine learning programmers will be needed.

Data Understanding: To achieve the goal, a variety of sources were used to gather data. First of all, data to analyze the consumption patterns will be needed. Thus, past data about the energy consumption per household [3][7][10][11] will set the basis to seek for a trend. Furthermore, the amount of carbon emission produced due to the energy consumption per household will be needed, in order to evaluate the effectiveness of using renewable energy. Moreover, looking into the relationship between energy consumption and temperature as well as the weather [1] could help set a finer trend. Secondly, data on the amount of renewable energy generated by the wind turbines will be needed to calculate the most efficient way to generate energy to reduce the financial loss. In regards to that, the location of the wind turbines[4][5] and their production rates [5] based on their locations will be needed to recommend future investments [6] in terms of location for renewable energy. Thus, wind [12] data of each location will also be needed.

Data Preparation: The tools required for this project are the programming language Python with its supporting modules: pandas, numpy, matplotlib, seaborn, scikit learn, and the Jupyter notebook to create and test machine learning models. There will be some additional costs accompanied for executing the project. For instance, a large dataset will need a database server and cloud gpu platforms to compute the model [13]. Moreover, some APIs will be charged for accessing the dataset.

The initial stage will be to link the respectively collected data into one dataset. For example, monthly energy consumption with monthly average of temperature from weather data. The next step is to clean the data by filling in missing data, checking and converting into uniformed data type, and normalizing or encoding the values. With this one dataset it will be possible to plot scatter plots, distribution plots, and heatmap to explore the features.

Regarding the General Data Protection Regulation(GDPR), the individual household energy consumption data may be an issue due to privacy, however, it is used to look at the general trend of energy consumption [14]. Thus, it might be the case that the individual energy consumption has to be processed into regional/district consumption data and handled with care in order to comply with the GDPR.

Modelling: To model the data, the first step will be to divide the dataset into a 7 to 3 ratio for training and testing set. In addition, the training set will be further split into a 7 to 3 ratio for training and validation sets for 5-fold cross validation in order to avoid overfitting of the model [15]**.**

With the *No Free Lunch Theorem* [16] in mind, the first machine learning algorithm which will be applied to this project is the multivariable regression. There are several other alternative regression algorithms suitable for this project (Appendix 1). Although the multivariable regression algorithm has drawbacks [17], our group has decided on it based on its benefit of helping one to understand the relationships among variables in the dataset if modeled correctly. The decision to change the regression algorithms will depend on how well the algorithm can be optimized. This will be the case, if the outcome, the energy consumption prediction, of the model accuracy will stay below 0.8(R-Squared) after the optimization [18].

Evaluation: The outcome of the model will give a predicted energy consumption and in comparison with the actual consumed electricity it will be possible to evaluate the accuracy of the prediction. For assessment of the business success criteria, financial profits gained or lost from generating renewable energy will be used. The denoted financial profit or loss measurements can be approached by observing the decrease in CO_2 emission by comparing its primary source of electricity production, burning fossil fuels, to the clean energy source.

Deployment: The results are model outcomes mentioned above and will be updated on a monthly basis. The stakeholders, preferably corporations in the renewable energy field, will be presented with a visualized whisker or a bar plot which depict the outcomes discussed in the evaluation part of this report. These graphs will assist in optimizing the business strategies by informing them how to generate more electricity with wind turbines in more productive conditions and reduce gas emission from utilizing alternative energy generation. If the project turns out to be successful, in the act of corporate social responsibility [19] or with the benefit of carbon offset credits[20] , there might be a repercussion that corporations will adopt the methodologies introduced in the project.

References

[1] Accuweather. (2021). accuweather. https://www.accuweather.com/en/nl/netherlands-weather

[2] CBS. (2018). Economy. Trends in the Netherlands 2018. Retrieved September 26, 2021, from https://longreads.cbs.nl/trends18-eng/economy/figures/energy/

[3] Easy Nuts. (2021, March 23). Expats set up utilities. https://easynuts.nl/set-up-utilities-iamexpat/

[4] Government of Netherlands. (2017, July 26). Wind energy on land. Information from the Government of The Netherlands Government.nl.
https://www.government.nl/topics/renewable-energy/wind-energy-on-land

[5] Government of the Netherlands. (2020, June 30). Offshore wind energy. Information from the Government of The Netherlands Government.nl.
https://www.government.nl/topics/renewable-energy/offshore-wind-energy

[6] Netherlands enterprise agency. (2021). Energy investment allowance - EIA RVO.nl. Netherlands Enterprise Agency RVO.nl. https://english.rvo.nl/subsidies-programmes/energy-investment-allowance-eia

[7] PartnerPete. (2021). Home Utilities in the Netherlands. Home utilities in the Netherlands | PartnerPete – Connecting Utilities. https://partnerpete.com/

[8] Rijksoverheid. (2021). Energie thuis. Informatie van de Rijksoverheid | Rijksoverheid.nl. https://www.rijksoverheid.nl/onderwerpen/energie-thuis

[9] State of the Planet. (2021, April 2). Social cost of carbon: What is it, and why do we need to calculate it? https://news.climate.columbia.edu/2021/04/01/social-cost-of-carbon/

[10] Vandebron. (2021). Nu tijdelijk tot €300,- korting op goeie energie. Vabdebron. https://welkom.vandebron.nl/affiliate/?utm_medium=affiliate&utm_source=daisycon&utm_campaign=daisycon

[11] Vattenfall. (2021). Energie: Stroom en gas. Energieleverancier Vattenfall – energie voor 2 miljoen klanten Vattenfall. https://www.vattenfall.nl/producten/energie/?cmp=AFC-740-381&affiliateid=259082

[12] Windy.app. (2021). Wind & weather forecast in Netherlands. WINDY.APP. https://windy.app/spots/Netherlands#?

[13] Geekflare. 2021. *10 Best Cloud GPU Platforms for AI and Massive Workload*. [online] Available at: <https://geekflare.com/best-cloud-gpu-platforms/> [Accessed 25 September 2021].

[14] Lexo Zardiashvili and Francien Dechesne, “Consumer Control of Energy Data: The Need for the Consent Management Mechanism in the Energy Sector of the Netherlands and Roadblocks Related to its Implementation”, 2019

[15] Google Developers. 2021. *Training and Test Sets: Splitting Data | Machine Learning Crash Course*. [online] Available at: <https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data?hl=ko> [Accessed 26 September 2021].

[16] Towards data science. 2021. *The Intuition Behind the No Free Lunch Theorem*. [online] Available at: <https://towardsdatascience.com/intuitions-behind-no-free-lunch-theorem-1d160f754513> [Accessed 26 September 2021].

[17] Sciencing. 2021. *The Difference Between Bivariate & Multivariate Analyses*. [online] Available at: <https://sciencing.com/difference-between-bivariate-multivariate-analyses-8667797.html> [Accessed 26 September 2021].

[18] Towards data science. 2021. *What are the best metrics to evaluate your regression model?*. [online] Available at: <https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b> [Accessed 26 September 2021].

[19] Triplepundit.com. 2021. *Why Renewable Energy Is Now a Requirement for CSR*. [online] Available at: <https://www.triplepundit.com/story/2016/why-renewable-energy-now-requirement-csr/23496> [Accessed 26 September 2021].

[20] RaboResearch - Economic Research. 2021. *Can voluntary carbon markets change the game for climate change?*. [online] Available at: <https://economics.rabobank.com/publications/2021/march/can-voluntary-carbon-markets-change-the-game-for-climate-change/> [Accessed 26 September 2021].

Appendix

[1] Alternative Machine Learning Algorithms based on the benefits and drawbacks

- Random forest regression
  - Benefits: Accurate, good performance on both linear and non-linear problems
  - Drawbacks: Not easily interpretable, overfitting can happen easily.
- Support vector regression
  - Benefits: Not biased by outliers, perform well on non-linear problems, easy to adapt.
  - Drawbacks: Outcome may be difficult to understand, feature scaling is essential.