|  |  |
| --- | --- |
| A picture of a winding road and trees  Carbon emissions: issuances vs retirements  Group 1 Interim Report | Interim report    Nahush Krishna: 0792816  Chinmay Chamoli: 0785561  Yashu Bhati: 0784198  Jenson Puthenpeedikayin Jacob: 0794547 |
|  |  |

Contents

[Abstract: 1](#_Toc132979630)

[Keywords: 1](#_Toc132979631)

[Introduction: 1](#_Toc132979632)

[Objectives: 1](#_Toc132979633)

[Related work 2](#_Toc132979634)

[Methodologies employed in modelling the dataset and results obtained: 2](#_Toc132979635)

[Data pre-processing: 2](#_Toc132979636)

[LSTM (Long Short-Term Memory): 6](#_Toc132979637)

[ARIMA: 7](#_Toc132979638)

[SARIMA: 7](#_Toc132979639)

[ARIMA and SARIMA are both forecasting algorithms. ARIMA considers past values (autoregressive, moving average) and forecasts future values based on them. SARIMA, like past values, takes seasonality trends into account. 8](#_Toc132979640)

[Results after optimization: 8](#_Toc132979641)

[Ethical Concerns: 10](#_Toc132979642)

[Consent: 10](#_Toc132979643)

[Clarity: 10](#_Toc132979644)

[Consistency and trust: 10](#_Toc132979645)

[Control and transparency: 10](#_Toc132979646)

[Consequences: 10](#_Toc132979647)

[Limitations: 10](#_Toc132979648)

[Conclusion: 11](#_Toc132979649)

[Plan for semester 4: 11](#_Toc132979650)

[References: 11](#_Toc132979651)

# Abstract:

Growing carbon dioxide levels in the environment is causing governments and organizations throughout the world to find effective ways to either reduce carbon emissions as much as possible or try to mitigate the damage caused by releasing greenhouse gases into the environment. One method that has gained popularity recently has been the carbon emissions – issuance vs retirements process. This project tries to analyse these 2 processes by obtaining raw data about companies looking for retirements and companies offering issuance, clean the data, analyse it, and finally help connect companies find the most cost effective and efficient matches to help them with retirements.

# Keywords:

Carbon footprints, Carbon emissions, Carbon retirements, Carbon issuances, Climate change, Greenhouse gas emissions, Sustainable practices.

# Introduction:

Carbon emissions: retirements vs issuances process has become the most crucial aspect of trying to battle climate change. This process allows companies to buy and sell credits to perform issuances or retirements so that the organizations or governments can reduce the impact of climate change that are caused due their greenhouse gas emissions.

This project is not considering all greenhouse gases but only issuances and retirements of carbon emissions. 1 credit is calculated as 1 metric tonne of CO2 that has been released into the atmosphere. Companies who incorporate long term processes to reduce their emissions are allowed to generate credits.

This means they get 1 credit for every metric tonne of carbon emission that they have managed to stop from being released into the atmosphere through their practices. The companies can then sell these credits to companies who are looking for retirements. This process encourages companies to find effective, sustainable ways to reduce emissions by providing financial incentives and reward the companies by helping them sell the credits that were generated by these processes.

This process is being used to generate a vast ecosystem where credits are constantly being generated and retirements are issued. The idea here is to reduce carbon emissions as much as possible. We will be using the datasets available to find patterns in among companies generating issuances and retirements and use the results to help companies for retirements find the right fit.

The most important outcome of using this process will be that companies will be incentivized to reduce carbon emissions and this process will in turn help reduce global carbon emissions to a large extent thus drastically slowing down if not eliminating it.

If carbon emissions are slowed down to a pace that nature can correct the issue then we can confidently state we will have a better, greener future for our next generation and the generations that come after.

# Objectives:

We would be predicting future Carbon emission issuance vs retirement trends based on past trends. The target of the analysis will be companies producing high CO2 emissions and companies generating credits due to implementing processes which have reduced their greenhouse gas emissions.

We intend to help organizations and companies analyze their offsets and help reduce their carbon footprints. We intend to forecast issuances vs retirements in the voluntary carbon market (VCM) using LSTM (Long Short-Term Memory) which is an RNN model.

Following this, we will also perform analyses using ARIMA and SARIMA models to perform a comparison between multiple models and choose the best model in the end based on the R-squared score. We hope that the analysis thus performed will help the organizations bridge the gap between the issuances vs retirements in voluntary carbon market.

# Related work

We have been interacting with an analyst working at a London based offsetting company called Allied Offsets. We were able to gain valuable experience on how the process works and the expectations and requirements of organizations. We received the VCM report which has been attached below.



In addition we have looked at various websites like [Carbon Direct](https://www.carbon-direct.com/insights/assessing-the-state-of-the-voluntary-carbon-market-in-2022) which provide an assessment of VCM and the various metrics used for calculation, the legal requirements and responsibilities, advantages and limitations of the VCM.

The [Berkley public policy](https://gspp.berkeley.edu/research-and-impact/centers/cepp/projects/berkeley-carbon-trading-project) website provides a list of links which contain various offsetting methods and how offsetting works in various industries and also the advantages and limitations associated with the particular industry.

# Methodologies employed in modelling the dataset and results obtained:

### Data pre-processing:

Initially, we had yearly carbon offset data for different projects. Since, the market is quite new, the time steps were very less so the predictions were not good enough.

We collected data again on a daily level and then grouped it by month. This gave us enough time steps to make some predictions, but the variance was still too high, and the model was unstable. At this time, our time series looked like this:

Chart, line chart

Description automatically generated

We checked the time series for stationarity using Ad fuller test and the p-value was more than 0.05. This indicated that our time series was not stationary and had to be treated for stationarity. Here are some of the graphs that helped us verify that the series was not stationary visually.

Chart

Description automatically generated with medium confidence

We could see the rolling mean as well as the rolling standard deviation changes with time. Just to be completely sure, we made a decomposition plot to break down the series into the components including trend, seasonality, and random noise.

Graphical user interface

Description automatically generated with low confidence

As we can see the series has a clear upward trend. To get rid of the upward trend, we must differentiate the time series with a period of 1. If the series also had a seasonal trend, we would have differentiated it with a period of 12 as well to negate that trend. We also scaled the data to treat the random noise that we can see in the residual component. In summary, we differentiated the time series with a period of 1 to get rid of the increasing trend and scaled the data to bring more stability to the models. After this, the time series looked like this:

Chart, line chart

Description automatically generated

Then we generated sequences from the monthly dates and split the data into train, test, and validation sets. Finally, we created the LSTM model and evaluated it using r-squared, but the results were not very impressive still. We then tried ARIMA and SARIMA model and the results were much better. ACF and PACF plots were used to select the p, d, q, and s parameters at this stage.

Here's how to use ACF and PACF plots for this purpose:

Autocorrelation plot (ACF): The ACF plot shows the correlation between a time series and its lagged values. A positive correlation at lag k indicates that the series is dependent on the value at that lag. The plot typically displays a slow decay in correlation over time for non-stationary series. The ACF plot is useful in determining the parameter q (number of lags to be included in the moving average component) of an ARIMA model. If the ACF plot shows a sharp drop-off after a certain lag, this suggests that the time series can be modelled using an MA(q) model.

Partial autocorrelation plot (PACF): The PACF plot shows the correlation between a time series and its lagged values while controlling for the effect of all shorter lags. A positive correlation at lag k indicates that the series is dependent on the value at that lag after controlling for shorter lags. The plot typically displays a sharp drop-off in correlation over time for non-stationary series. The PACF plot is useful in determining the parameters p (number of lags to be included in the autoregressive component) and d (the degree of differencing required to make the series stationary) of an ARIMA model. If the PACF plot shows a sharp drop-off after a certain lag, this suggests that the time series can be modelled using an AR(p) model, where p is the lag after which the plot drops off.

Interpretation: Once you have plotted both the ACF and PACF, you can use them to determine the values of p, d, and q. If both plots show a sharp drop-off after a certain lag, this suggests that the time series can be modelled using an ARIMA(p,d,q) model, where p, d, and q are the respective lags after which the plots drop off. If the ACF plot shows a slow decay, this suggests that the series may require differencing (d) to make it stationary. If the PACF plot shows a slow decay, this suggests that the series may require an autoregressive (AR) component (p). If both plots show slow decay, then it may be necessary to include both an autoregressive and moving average component in the model (ARMA(p,q)).

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated

The results have been documented below:

### LSTM (Long Short-Term Memory):

LSTM is an abbreviation for long short-term memory networks, which are utilised in Deep Learning. It is a kind of recurrent neural networks (RNNs) that may learn long-term dependencies, particularly in sequence prediction tasks.We started off with trying to predict the future trends in the carbon offset market based on the past trends. We chose R-squared as the evaluation metrics of our model. The reason for choosing r-squared was the large values present in the data as well as the high variance of the dataset. We wanted to check how much of the variance could be explained by out machine learning models. We got the following results on using the LSTM model.

Chart, line chart

Description automatically generated

**R-squared - (-0.2009134531875536)**

The LSTM model gave us a negative R-squared value and the result is a poorly fit model as shown from the LSTM visualization. This was a major surprise since we expected that we would at least get a positive R-squared value. We tried to further analyse the reason and figured out that this model was not the best option for performing our analysis.

## ARIMA:

ARIMA, or autoregressive integrated moving average, is a statistical analysis model that uses time series data to better understand the data set or anticipate future trends. Autoregressive statistical models anticipate future values based on past values.

Chart, line chart, histogram

Description automatically generated

**R-squared value - 0.5644968819947727**

### SARIMA:

Seasonal Autoregressive Integrated Moving Average, or Seasonal ARIMA, is a modification of ARIMA that explicitly handles univariate time series data with a seasonal component.

Chart, line chart

Description automatically generated

**R-squared - 0.5822483440846893**

As we can see, ARIMA and SARIMA are able to capture the trends of the carbon offset markets. The variance of the dataset is high, and SARIMA can explain 58.22% of it.

We believe that the remaining variance is due to randomness of the nature of the voluntary carbon markets and the fact that it is still a developing market.

For sanity check, we also used Tableau's forecast functionality and to our surprise their model gave a negative value (-52.79940205545258) of R-squared just like our LSTM model. This suggests that our dataset is hard to model and ARIMA and SARIMA are the best choices.

After these initial results, the data still had a lot of variances, and our models were affected by it so we decided to explore it further. We treated the data for outliers to bring more stability to our models and the r-squared increased further. Finally, we did hyper-parameter tuning for ARIMA and SARIMA as they had the maximum potential and got the following results.

# ARIMA and SARIMA are both forecasting algorithms. ARIMA considers past values (autoregressive, moving average) and forecasts future values based on them. SARIMA, like past values, takes seasonality trends into account.

# Results after optimization:

We tried to optimize our models to give better R-squared values by trying feature engineering and we were able to get the following results. We have achieved the following results.

Best LSTM optimization:

Chart, line chart

Description automatically generated

**R – squared value: 0.31**

Best ARIMA optimization:

Chart, line chart

Description automatically generated

**R – squared: 0.65**

Best SARIMA optimization:

Chart, line chart

Description automatically generated

**R- squared: 0.73**

# Ethical Concerns:

### Consent:

The dataset that we used is a combination of individual datasets from multiple sources. We ensured that we had all relevant permission for the use of the datasets. We ensured that all the datasets we picked were for public use and that all access permissions were provided before we downloaded and used the dataset.

### Clarity:

We obtained the dataset from websites which had provided datasets for public use and analysis. We ensured that the datasets were used purely for research purposes and analysis.

### Consistency and trust:

We will be using the datasets for research and analysis purposes only and we will ensure that the data will not be sold or distributed to any third party for any reason whatsoever. We will be using the datasets to create unbiased analysis and ensure that our findings are not leaning towards any side (organizations looking for retirements or organizations providing issuances).

### Control and transparency:

The results obtained from the analysis will be strictly used for academic purposes only. He results will be used to analyse past issuances vs retirement trends and predict future trends. We will not be publishing or distributing the results on any external websites, and we will ensure that none of the findings cause any conflict or disagreements between any organizations or governments.

### Consequences:

The best model can explain around 74% of the variance of the time series data but 26% variance is random. If we give a prediction and the prediction is far from the actual carbon credits, then the client companies can incur losses which is not a good scenario. The model can create profits for our clients, but it might also bring in some losses sometimes. This might have financial consequences for our project as well as our stakeholders. To reduce the impact of these consequences, instead of predicting a number we should use statistics to determine the confidence intervals based on our predictions. By employing statistics, our predictions will be way more reliable, and the consequences will be minimal.

Other than the financial consequences, the data is not sensitive so there are no consequences to the data breaches. Once we start collecting client data, we need to take appropriate steps to safeguard the data and protect it from any breaches.

# Limitations:

Due to very large observations in the dataset the variance observed was very high which caused issues with model stability.

We tried our best to include as many carbon registries as we could, but it is not possible to cover them all due to lack of time and lack of transparency in VCM (Voluntary Carbon Markets) transactions.

VCMs are an emerging market and we do not have enough research topics to understand the markets better. Due to this we are doing a research-based analysis.

# Conclusion:

Successfully performing analysis on VCM will ensure that the companies looking to retire carbon offsets find the best organizations to help them out and vice versa. In addition, the analysis of the past trends will help improve analysis and prediction of future trends and the analysis will help organizations find the best fit for their requirements. Our intention is to become the ‘middleman’ between companies looking for retirements and the companies looking for issuances. Since VCM is an upcoming market there is a massive potential for growth in terms of generating revenue, helping reduce carbon emissions and opens a huge door for performing data analysis on different markets and approaches. We feel that what we have analysed is just the tip of the iceberg and that this is just the beginning. There is a vast world of analysis open in this market, we can try to analyse other greenhouse gases and create further analysis which can help gain valuable insights and can help the VCM reach new heights.

# Plan for semester 4:

We have done the analysis for Issuances and retirements in LSTM. We have completed analysis in ARIMA and SARIMA for issuances only. We will be working on analysis of retirements using ARIMA and SARIMA in the next sem. In addition, we intend to try and get more data for analysis post with if we have enough time, we will create a web application in which we will try to find organizations actively looking for retirements and connect them with the right partners.

# References:

Author: Carbon Direct,

Title: Assessing the State of the Voluntary Carbon Market in 2022,

Source: Carbon Direct Insights,

Date created: 2022-01-25,

URL: <https://www.carbon-direct.com/insights/assessing-the-state-of-the-voluntary-carbon-market-in-2022>

Centre for Environmental Public Policy. (n.d.). Berkeley Carbon Trading Project. University of California, Berkeley. Retrieved April 13, 2023, from <https://gspp.berkeley.edu/research-and-impact/centers/cepp/projects/berkeley-carbon-trading-project>

Dataset links:

Verified Carbon Standard. (n.d.). All Projects. <https://registry.verra.org/app/search/VCS/All%20Projects>

Gold Standard. (n.d.). Gold Standard Projects. <https://registry.goldstandard.org/projects?q=&page=1>

Gold Standard. (n.d.). Gold Standard Credit Blocks. <https://registry.goldstandard.org/credit-blocks?q=&page=1>

APX. (n.d.). Market Reports. <https://acr2.apx.com/myModule/rpt/myrpt.asp?r=111>

APX. (n.d.). Emission Reduction Credit Reports. <https://acr2.apx.com/myModule/rpt/myrpt.asp?r=112>

APX. (n.d.). The Reserve. <https://thereserve2.apx.com/mymodule/mypage.asp>

Markit. (n.d.). Project Registry - Public View. <https://mer.markit.com/br-reg/public/index.jsp?entity=project&srd=false&sort=project_name&dir=ASC&start=0&entity_domain=Markit&additionalCertificationId=&acronym=PV&standardId=100000000000004&categoryId=100000000000001>

Global Carbon Council. (n.d.). Approved Projects. <https://projects.globalcarboncouncil.com/pages/approved_projects>

Climate Forward. (n.d.). Home. <https://climateforward.apx.com/>

Research Links:

<https://www.greenbiz.com/article/carbon-offset-prices-set-increase-tenfold-2030#:~:text=Carbon%20offset%20prices%20on%20average,based%20solution%20projects%2C%20and%20provide>

<https://www.globenewswire.com/en/news-release/2022/10/28/2543711/0/en/Voluntary-Carbon-Offsets-Market-Size-2022-2027-is-Projected-to-Reach-USD-700-5-Million-at-a-CAGR-of-12-71-Industry-Share-Demand-Growth-Rate-Demand-Revenue-Key-Players-Market-Status.html#:~:text=global%20Voluntary%20Carbon%20Offsets%20market,USD%20700.5%20million%20by%202027>

<https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions-data>

<https://www.frontiersin.org/articles/10.3389/fenvs.2021.721517/full>

<https://aegis-hedging.com/insights/voluntary-carbon-credit-registry-issuances-and-retirements-january-april-2021-vs-2022#:~:text=The%20total%20number%20of%20issuances,same%20four%2Dmonth%20comparative%20period>

<https://www.ecosystemmarketplace.com/publications/closing-the-carbon-offsets-issuances-retirements-gap/>

<https://www.carbonneutralindiana.org/what-does-it-mean-to-retire-carbon-offsets>