Spring 2019

Climate Change Analytics: Predicting Carbon Price and CO2 Emissions

Edward Davis

Climate Change Analytics:

Predicting Carbon Price and CO_2 Emissions

Edward Davis

La Salle University, Philadelphia, PA 19141 USA

TABLE OF CONTENTS

ABSTRACT

The focus of this research is to predict the greenhouse gas emissions and the funding to help combat this global problem. There must be consistent funding to support and sustain the planet ecosystems. This research is motivated by the global concern of climate change caused by greenhouse gas emissions and the need to consider a multinational strategy to provide funding to combat it. The goal of the funding is to provide adequate financial backing and support for innovations needed to combat this problem.

This research leverages the capabilities of machine learning found in Weka and forecasting and visualization in Tableau. The models are expected to predict a carbon tax rate that could be used multi-nationally. The results and performance measures will be scrutinized to identify the model that is the best fit for the proposed solution. The economic, population, land temperature, current multinational carbon tax rates and reverse carbon initiatives data will be interrogated by supervised machine learning models or classifiers (Frank et al., 2011). The CO_2 emissions for China, India and the United States will also be predicted to show expected increases in emission based on historical data through Tableau forecasting.

This study concluded that a carbon rate can adequately be created and predicted using machine learning models. And, CO_2 emissions can also be predicted using public open data sources that provide economic, population and surface temperature features.

INTRODUCTION

This research is motivated by the environmental problem of greenhouse gas emissions. The research presents analysis and models for various aspects of this problem. The effects of greenhouse gas emissions' criticality, and impact of population growth will be reviewed to show carbon tax can be predicted using machine learning (McNall, 2012).

The Intergovernmental Panel on Climate Change (IPCC) emphasized the need to establish a tax on CO_2 emissions as an instrumental mitigation tool. A carbon tax directly sets a price on carbon by defining a tax rate on greenhouse gas emissions (Global warming of 1.5[degrees]C, an IPCC special report, 2019). A carbon tax sends a price "signal" through the economy to get energy companies and startups to ramp up on low-carbon investments and search for reduction strategies for CO_2 emissions.

Many scientists are worried about an increase in global warming to 2^0C which is caused by Greenhouse gases (Cote, 2019). Since the top emitters of greenhouse gases happen to be several of the world's largest countries and alliances (i.e., China, India, United States, and the European Union), it is expected that this problem should get the appropriate level of priority and urgency it deserves (Global warming of 1.5[degrees]C, an IPCC special report, 2019). This work looks at emissions from three of the largest countries: United States, China, and India.

The year 2050 is the target on some timelines when the world's powers, countries, concern stewards, scientists, and stakeholders will be checking on how successful they have been progressing against the greenhouse effect. They will also look at what carbon reversal solutions can be implemented to help the world become Carbon Neutral (What is Carbon

Pricing? n.d.). This work includes Tableau predictions of population, emissions, and surface temperatures through 2052.

The focus of this study is to use machine learning in Weka to produce relevant Carbon tax predictions based on features/attributes[1] like Carbon Price Initiatives, the human development of countries, initiative-related cost factors and relevant economic indicators. The study will also use machine learning in Weka and linear forecasting in Tableau to produce adequate CO_2 predictions based on economic, population, and surface temperature data features.

The study will pursue answers to the following:

- Can a Carbon Tax be predicted based on its relationship to Carbon Price Initiatives Value and Countries' Human Development Rank? A carbon tax is seen by many as an essential part of the solution. There are many ways out of this dilemma, but it means changing what is done and how to do it. (McNall, 2012)
- Can CO_2 emissions be predicted based on its relationship to population, consumption, surface temperature and other relevant economic indicators?

Scientists agree that humans are the blame for a good fraction of the planet's warming. A tax on carbon helps place the burden back on those who are responsible for the pollution or emission. By the end of the century (2100), it is expected that the planet can be resident to 11 billion people. This paper includes analysis projecting population growth, as well as growth in CO_2 emissions and related measures. David Satterthwaite (international

[1] Note, features, attributes, and database columns are synonyms.

institute of environment and development (UK), has stated, "Changes in our consumption are the key drivers of global warming more so than increasing the number of people on the planet (Satterthwaite, 2009). Higher consumption is what drives anthropogenic climate change, or the production of greenhouse gases emitted by human activity" (McNall, 2012).

LITERATURE REVIEW

The literature section will review machine learning techniques and the carbon price predictions research articles. Note, there are two main types of carbon pricing: emissions trading systems (ETS frequently referred to as "cap and trade") and carbon taxes. Each research study explores a different aspect of carbon price forecasting using machine learning and forecasting a price assigned to CO_2 emissions. ETS caps the total level of greenhouse gas emissions and allows those industries with low emissions to sell their extra allowances to larger emitters. A carbon tax directly sets a price on carbon by defining a tax rate on greenhouse gas emissions or – more commonly – on the carbon content of fossil fuels. The articles provide significant insights into the area of carbon pricing that establishes a price for purpose of charging for CO_2 emitted, each using the application of machine learning methods to derive the cost of CO_2 pollution.

A. Price Forecasting for Carbon Credits

The goal of this research article is to show the drivers behind the changes in price of carbon credits in the European Union Emission Trading Scheme (EU ETS). The study explains machine learning approaches used for the research work. The team chose to focus on neural network algorithms for prediction since the United Kingdom (UK) energy data is categorical, rather than continuous.

In response to the Kyoto Protocol, the European Union (EU) began preparing an EU carbon market. The European Union Emission Trading Scheme (EU ETS), the first international

cap and trade trading system was designed to establish overall emission levels or caps and enable EU members the capability to freely buy or sell emission allowances. The goal is to help EU Member States to meet their commitments to CO_2 reduction in a cost-effective way. European Union Allowance (EUA) price predictions are made from data provided by the UK energy market and equity markets (Guðbrandsdóttir & Haraldsson, 2011). The study presents a detailed description of modeling techniques used to predict EUA prices leveraging machine learning techniques.

There are limited analyses done that focus on UK energy data, even though the United Kingdom is the second largest emitter of EU countries that participate in the EU ETS. Certified emission reduction units (CERs) are leveraged to show same-day market relationship and can be a good predictor of EUA prices.

Certified emission reduction units were determined to be the only feature whose adjusted p-value was within the confidence interval of 95% (p-value below 0.05). CER had a strong same-day relationship with EUA returns and the model captures over 80% of the variability of EUAs.

There are several key takeaways from this study's work:

- There is a range of new market data such as the European Union Allowance (EUA) price that can provide interesting results or relationships never examined for CO_2 emissions.
- Linear regression can be a good machine learning technique or tool used to predicted continuous variables such as CO_2 emissions.
- The scope of the research study is aligned with one of the research goals of this capstone paper, however this study focuses on Cap and Trade instead of a Carbon Tax approach.

B. Carbon price forecasting using hybrid modeling

The focus of this research article is twofold: first, producing accurately predicted carbon prices leveraging machine learning practices and secondly, establishing a hybrid methodology that helps fills a gap when predicting carbon prices whose input data consists of both linear and nonlinear patterns (Zhu & Wei, 2013).

The authors, Zhu & Wei, chose two machine learning algorithms for their study and together the algorithms create the hybrid methodology. The autoregressive integrated moving average (ARIMA) model has been found to be one of the most popular models for predicting time series data because of its statistical features. Careful consideration was given when the ARIMA model was selected for this study. Primarily because, it is a class of linear model that just captures linear patterns in a time series and cannot capture nonlinear patterns hidden in the same time series.

The least squares support vector machine (LSSVM) was selected by the authors to complement ARIMA because it can solve linear problems quicker with a more straight-forward approach. Until now, LSSVM has been successfully used in pattern recognition and nonlinear regression estimation problems. This hybrid methodology decomposes carbon prices via these two components: a linear component and a nonlinear component.

The European Climate Exchange (ECX) located in London, is the largest carbon market under the EU ETS and tracks a great number of carbon prices. The authors chose, as experimental samples, the two main carbon future prices that mature in December 2010 (DEC10) and December 2012 (DEC12). These two carbon price indicators are the most famous benchmark prices and have traded on the market since the opening of EU ETS in April 2005. The data for the two carbon prices used are updated daily and freely available from the ECX website (http://www.theice.com).

In this study three hybrid models were used where a nonlinear regression function is determined by the LSSVM model and linear is determined by the ARIMA model. The best model achieved superior forecasting performances and produced good prediction results. This model is well suited for prediction with highly nonlinear and complex carbon price data. It proved to be a very promising methodology for carbon price forecasting.

There are several important takeaways from this research:

- The article reinforces the need for good carbon price forecast models whether they be single or hybrid model (i.e., combining linear and nonlinear models to create a hybrid package).
- The authors decided to leverage more than one tool or model with different strengths to create a solution to a complex problem.
- The authors' foresight created a solution that will fill a gap and enable future efficiencies for predicting carbon prices.
- The authors acknowledged that there is very little literature regarding forecasting carbon price.
- Both Carbon Price types: carbon tax and cap & trade represent the cost of CO_2 pollution.
- Carbon tax changes over time and can be treated as a time series process.

C. Forecasting of Energy-Related CO 2 Emissions in China for Sustainability

The goal of this research article (Dai, Niu, & Han, 2018) is to create an accurate forecast of CO_2 emissions for China with consideration given to its population. This forecasting will assist with China's CO_2 emission reduction policy.

The authors chose two machine learning algorithms for their study and together they create a hybrid methodology. The model used to forecast the main influencing factors of CO_2 emissions is Grey Model (GM) (Ye, Xie, Zhang, & Hu, 2018) (see APPENDIX A). Next, the least squares support vector machine (LSSVM) optimized by the modified shuffled frog leaping algorithm (MSFLA) (MSFLA-LSSVM) model is used to forecast the CO_2 emissions from the relevant input features.

The forecasting accuracy of CO_2 emissions is affected by many factors. The influencing factors interrogated are population, carbon emissions intensity, GDP, total coal consumption, urbanization rate, industrial structure, energy consumption structure, energy intensity, total imports and exports and other factors of CO_2 emissions (Dai, Niu, & Han, 2018). Feature dimension reduction was used to identify and chose the CO_2 emissions forecasting model's input features. They are per capita GDP, urbanization rate, total coal consumption and total imports and exports.

Empirical analysis is conducted, and it verified that the MSFLA-LSSVM model has strong generalization ability and the robustness for CO_2 emission forecasting. The analysis also determines that the forecasting accuracy of MSFLA-LSSVM is better than that of previous machine learning and neural network models. It is superior in performance and a better choice for CO_2 emissions forecasting.

There are several important takeaways from this research article:

- The study and documentation provided a good reference for predicting CO_2 emissions presented in this paper's study.
- Nonlinear data can be leveraged to improve the accuracy of the predicted CO_2 emissions. Also, the idea of combining models to derive the best solution set should always be considered.
- Data preparations capabilities, like the grey relational degrees, derive feature reduction which provides more information to support better prediction performance and reduces input file size.

The key point from this literature review is that all three research articles discuss the importance of predicting carbon prices (tax or cap & trade) and emissions to help support crucial global CO_2 emission reduction initiatives. This supports the response to global climate change.

RESEARCH

The purpose of this research is to analyze data related to various aspects of climate change. This involves forecasting using linear regression, as well as predictions using machine learning. Machine learning methods are used for creating carbon tax rate predictions as well as creating CO_2 emissions predictions. Linear regression is done on several relevant attributes.

The research methodology used to accomplish the work for this study consists of model development using data provided by reliable sources and updating missing data to construct complete datasets. Models were developed in Weka and Tableau. The learning algorithms used in Weka are ZeroR, DecisionStump, REPTree and RANDTree. The models created were put through a comparison to determine which one offers the best carbon tax rate option. Tableau forecasting and visualizations were plotted against World Population, World CO_2 Emissions (Gt CO_2), China Final consumption expenditure, India Final consumption expenditure, US Final consumption expenditure, and Global Means Surface Temperature Change (12 Month Avg data points). Forecast or future projections were made to show increases in population, CO_2 emissions, final consumption expenditures (China, India, and US) and surface/land temperature change. The research framework is shown in Figure 1 Research Framework. The framework defines the steps in the machine learning research process for this study.

Figure 1 Research Framework

Research Approach

Select Data Sources

The data for this study was obtained from the World Bank (WB) website (World Bank Open Data, 2015) associated with the United Nations (UN), Worldmeters Population by Year website, and Drawdown.org website (Drawdown: 100 Solutions to Reverse Global Warming, 2019). The selected data includes multinational data sources:

- World Bank Economic Data
- World Bank Population Data
- World Bank Carbon Price Data
- Land Temperature Data
- Drawdown.org Reverse Global Warming Solutions

Preliminary Data Analysis

Datasets Used:

Dataset 1 contain the World Bank Carbon Price, World Bank Human development index (HDI) and Drawdown.org: Reverse Global Warming Solutions data (Drawdown.org, 2019) that used to predict carbon tax. The data set has 65 instances of initiative data and class value.

Datasets 2 to 4 contain the Country [equals US (2), China (3), and India (4)] Economic Indicators, GHG emissions, and Land Temperature data.

The data set has 59 instances of Country data and the class value predicted. There were 21 attributes for each of the 3 countries. Attribute selection (below) reduced the number of attributes to be used for prediction.

Address Missing Data

It is common during machine learning data preparation to have missing values. The features in this study are mainly numeric. It common practice to replace missing values with the average value for the feature. This was applied to the predictor features.

Instead of the average value, the carbon price feature was replaced by the United States Environmental Protection Agency (EPA) social cost of carbon priced at $42 for year 2020 as a substitute for missing or blank data entries (The Social Cost of Carbon, 2017). Currently, the United States does not have a federal or state level carbon tax. The EPA and other federal agencies use estimates of the social cost of carbon ($SC\text{-}CO_2$) to value the climate impacts for rulemaking that support guidance like car and truck emission standards. The social cost of carbon is an economic measure expressed as the dollar value of the total damages from emitting one ton of carbon dioxide into the atmosphere. The EPA's social cost of carbon has considered and incorporate factors that would be included in a carbon tax (The Social Cost of Carbon, 2017). The social cost of carbon is comparable to a carbon tax because both set a price or cost for damages from emitting CO2 into the atmosphere. This measure is like the British Columbia Carbon Tax (Carbon Pricing Dashboard (n.d.)). The use of the cost of Carbon skews the results toward science versus political concerns.

Terms

Machine learning numeric predictions can be evaluated using a variety of measures, including accuracy, correlation, mean absolute error, root mean squared error and root relative squared error. These are explained below.

Accuracy: looks at the proportion of a complete sample set that makes up the total number of predictions determined to be correct.

Correlation coefficient (r) is a measure of the predictions and actual/target values association. It measures the strength in the linear relationship. Correlation is a statistical technique for measuring the strength of the relationship between two variables (e.g., age and blood sugar). When a correlation is perfect, it is 1.0, but that does not mean that the predictions are perfect. A value close to 1.0 or -1.0 is good given the scale -1.0 – +1.0. For example, a correlation coefficient of 0.2055 implies 20.55% of the variance in the data is explained by the model. Note, a low value isn't bad if it is the best model fit.

MAE (Mean Absolute Error): $MAE = f(x) = \frac{1}{n} + \sum_{n=1}^{\infty} |Predicted_n - Actual_n|$

MAE is the mean of the absolute errors. The absolute error is the absolute value of the difference between the forecasted/predicted value and the actual value. The MAE shows how big of an error can be expected from the forecast. It also produces the average magnitude of the errors for a set of predictions

RMSE (Root Mean Squared Error): $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_{i} - Actual_{i})^2}{n}}$

RMSE is a measure of accuracy which is used to compare forecasting errors of a particular dataset. It is always non-negative and a value of 0 indicates a perfect fit to the data. A lower RMSE is better than a higher one. Lower values of RMSE indicate better fit.

If the RMSE and MAE are the same, then all the errors are of the same size and importance. Comparing RMSE and MAE can be used to determine whether the forecast contains large but infrequent errors. RMSE gives large errors greater importance since the errors are squared. If the RMSE is significantly larger than MAE that is a sign that the error size is inconsistent, with large errors contributing to the large value.

RRSE (Root relative squared error): $RRSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_{i} - Actual_{i})^2}{\sum_{i=1}^{n}(Actual_{i} - \overline{Actual})^2}}$

RRSE is computed by dividing the RMSE by the RMSE obtained by just predicting the mean of target values (and then multiplying by 100). So, the smaller values are considered better fit and values > 100% indicates a scenario that is doing worse than just predicting the mean or average. Better models are usually those with accuracy as high as 99%.

AIC (Akaike information criterion): $= n * \log(SSE/n) + 2 * (k+1)$

AIC is a model quality measure, developed by Hirotugu Akaike, that penalizes complex models to prevent overfitting. The model fit is by maximum likelihood and the lowest AIC identifies the better choice. In this definition, k is the number of estimated parameters, including initial states, and SSE is the sum of the squared errors. This will be used to evaluate Tableau predictions.

Analytical Methods

This study will apply machine learning using Weka and forecasting in Tableau. The machine learning algorithms used will build a mathematical tree based on sample data (or training data) to make predictions or decisions without being programmed to perform the expected task. Some graphical visualizations and time-series forecasting will be conducted in Tableau.

Tableau Forecasts: Population, Consumption, Surface Temperature and Carbon (CO_2) Emissions

Can increases in population growth impact greenhouse emission? There are some concerns to be confronted when involving human population growth, behavior and activities conducted on a normal daily basis. These conditions become more apparent as seen in the linear trends and forecasts in the analysis that follows. Many of our normal daily activities such as burning fossil fuels (coal, oil and natural gas), agriculture and land clearings increase the concentrations of greenhouse gases emitted back into the atmosphere (EPA: Greenhouse Gases, 2017).

The world's population growth will require more consumption of the Earth's natural resources and will contribute to the problem. The places and resources that provide for current consumption are likely to be the only resources to provide for the additional population growth. CO_2 taxation will support mitigation programs that will find innovative ways to provide sustainable water and energy resources to accommodate the increase in global populations.

Figure 2, was Excel generated from historical data provided by the World Bank Group. It shows annual CO_2 emissions predictions for China, India, the United States, the European Union and the rest of the world from 1970-2017. This illustrates the growth in CO_2 emissions throughout the world. It's shown particularly in China over the last 15 years and "the

rest of the world" over a slightly longer timeframe. Emissions in the United States and the European Union actually show a decrease. A potential explanation is that many businesses in the US and EU are making conscious effort to reduce CO2 by adopting CAP and Trade practices, using less coal or switching from coal to other sustainable energy sources such as wind and solar, and using methods like CO2 capture and storage. (Lackner, et al., 2012)

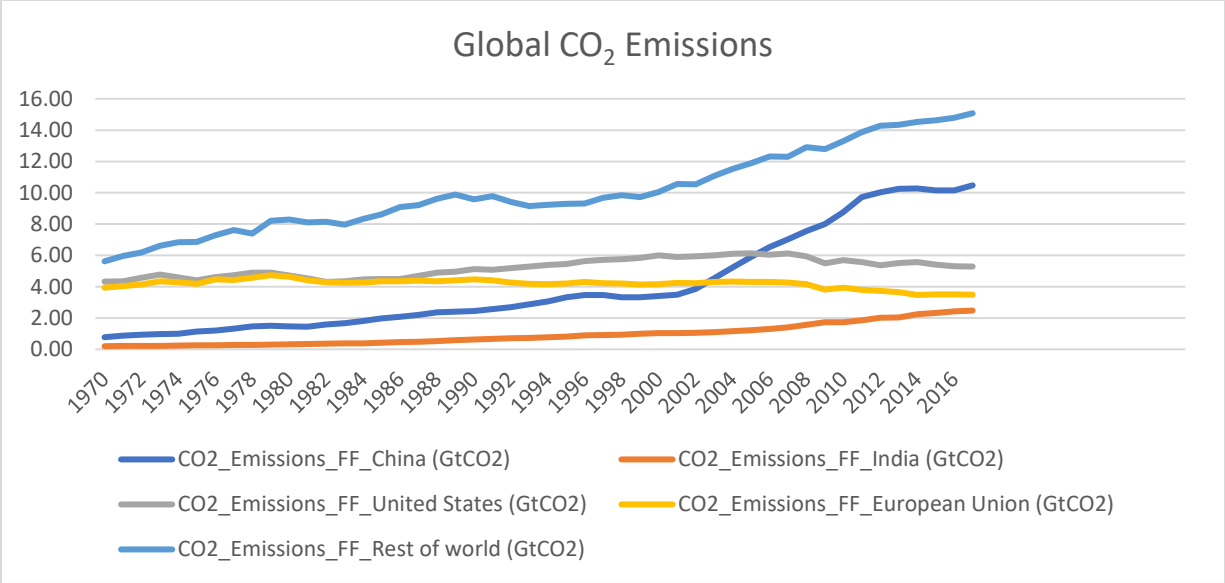

Figure 2: Change in Emissions 1970-2017

Further analysis suggests relationships between different features. **Figure 3** shows Tableau generated graphs of World Population, China Final consumption expenditure, US Final consumption expenditure, and India Final consumption expenditure. Each feature is plotted against World CO_2 Emissions (Gt CO_2) and shows a continuing upward trend suggesting potential strength in the relationship with emissions.

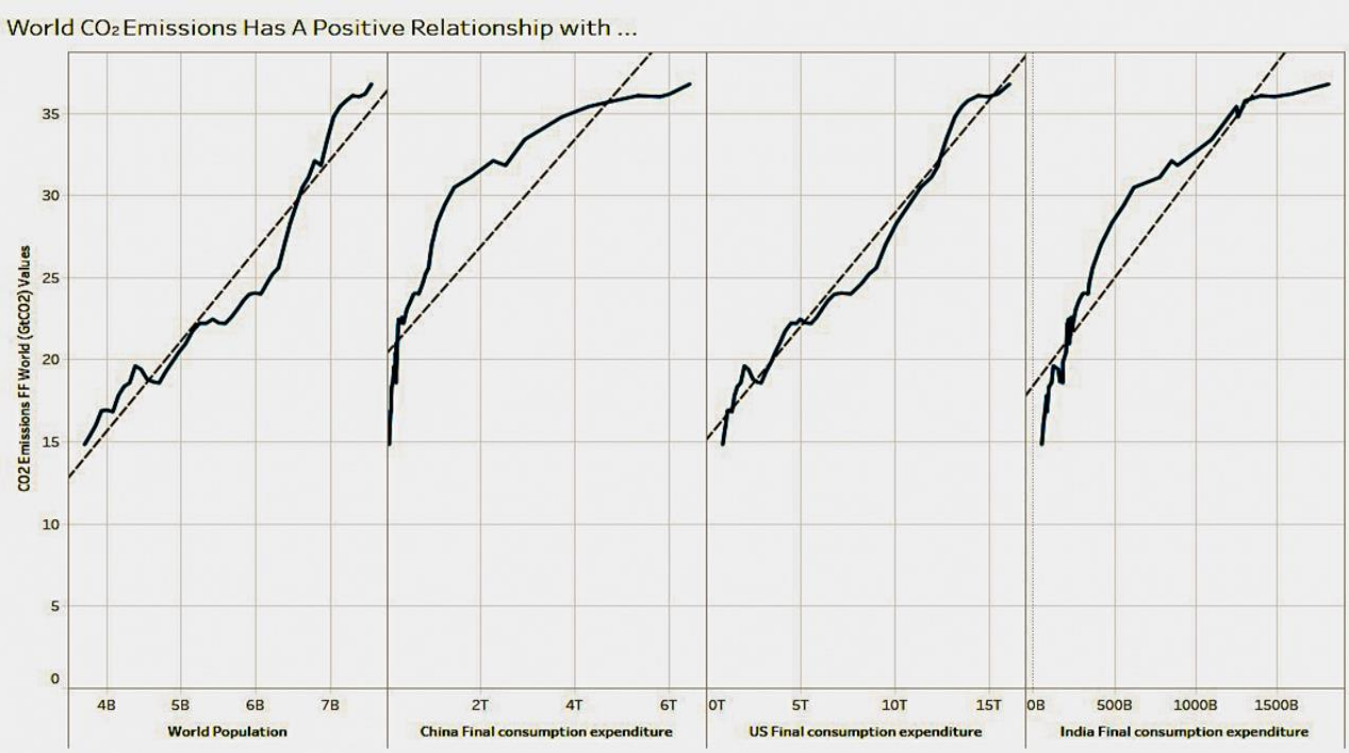

Figure 3 Consumption and population correlation with Total World CO_2 Emissions

Figure 4 shows positive linear movement for population, three economic features, CO_2 emissions, and temperature. This positive movement is also realized in the forecast. Each forecast model was created in Tableau and used the same input features as those used in the Weka machine learning models.

Each feature is shown in relationship with CO_2 emission and shows similar trends. From the visualization, it is believed population affects economic activity (i.e., consumption expenditures) and consumption expenditures affect CO_2 emissions. It's also a belief that CO_2 emissions impacts temperature.

Because the graphs show a strong positive linear relationship between the World CO_2 Emissions and the selected input features, it can be hypothesized that these features will be good

predictors of World CO_2 Emissions. The graphs also show predictions of future trends on these features.

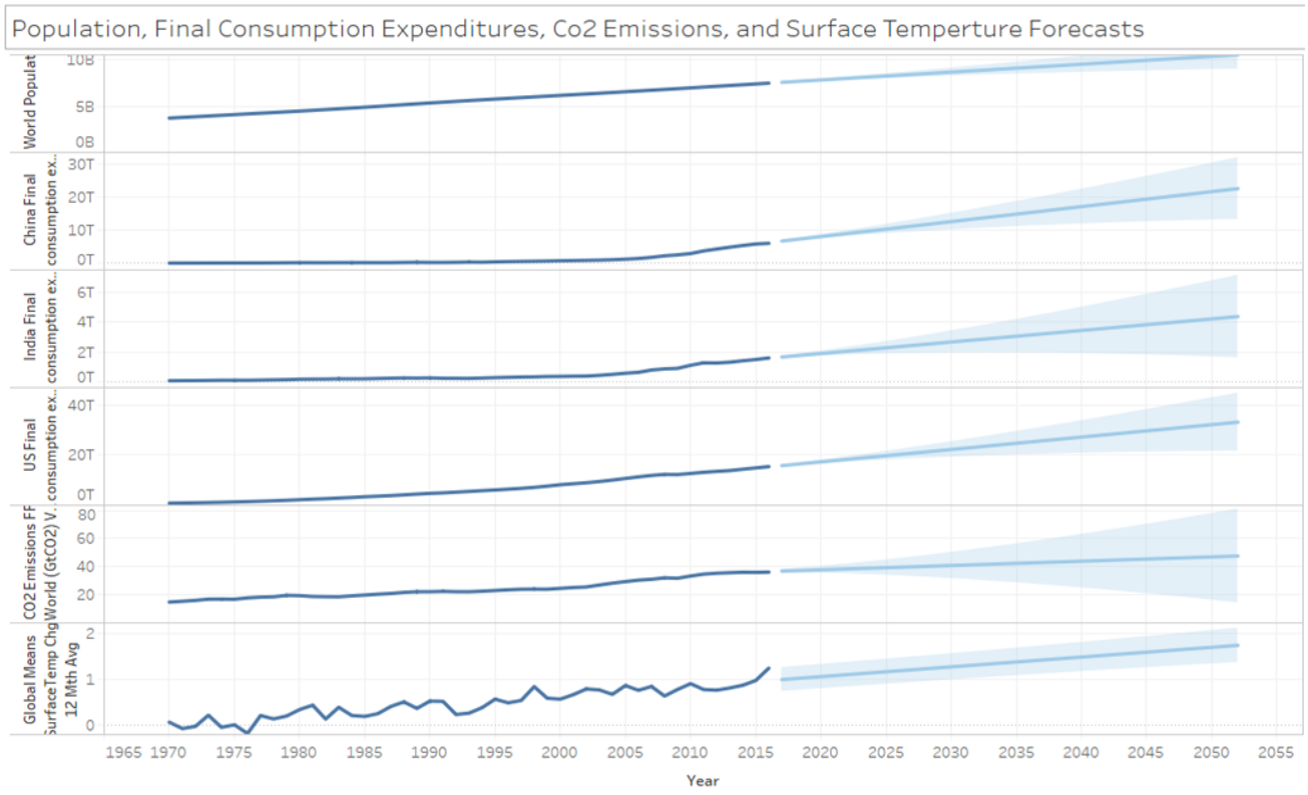

Figure 4 – Trendlines with predictions for population, economic output, emissions, and temperature

Table 1 shows the 2017 World Population, China Final consumption expenditure, India Final consumption expenditure, US Final consumption expenditure, World CO_2 Emissions (Gt CO_2) and Global Means SurfaceTemp Chg data values along with the derived change values that get to the 2052 forecast values. All forecasts were computed using Tableau. From the forecasts, it can be concluded that by 2052 world population will be approximately 10.5 billion people. Adding almost three (3) more billion people who become new consumers of the planet's resources should in some way impact global warming. From a basic economic perspective, supply and demand will be impacted by the additional people.

China's final consumptions expenditures will be near $22.5 trillion, India's will be near $4.4 trillion, and the US will be near $33 trillion. The World CO_2 emissions will be close to 48 (Gt CO_2) and mean surface temperature will increase by nearly 1.7^0 C.

Table 1 Tableau Linear Forecast Summary

Features	**Initial** **2017**			**Change from Initial** **2017 – 2052**	**Forecast**
World Population	7,552,397,525	±	66,444,801	2,972,411,862	10,524,809,387
China Final consumption expenditure	6,685,697,503,215	±	279,934,544,264	15,790,321,812,886	22,476,019,316,101
India Final consumption expenditure	1,652,922,341,623	±	83,936,120,947	2,731,063,124,305	4,383,985,465,928
US Final consumption expenditure	15,824,673,990,295	±	348,158,987,601	17,346,268,154,981	33,170,942,145,276
CO_2 Emissions FF World (Gt CO_2)	36.92	±	1.34	10.76	47.68
Global Means SurfaceTemp Chg 12 Mth Avg	0.991	±	0.258	0.740	1.731

Table 2 **Tableau Linear Forecast Evaluation Summary** shows statistics and indicators that describe the accuracy and quality of the model ran for forecasting in Tableau. Note, the Quality column indicates how well the forecast fits the actual data. Possible values are GOOD, OK, and POOR. Quality is expressed relative to a naïve forecast, such that OK means the forecast is likely to have less error than a naïve forecast, GOOD means that the forecast has less than half as much error, and POOR means that the forecast has more error.

The lowest AIC is used to identify the model with the best fit for estimating the likelihood of predicting/estimating future values (Chan & Tsay, 2012). RMSE and MAE are viewed as well. If the RMSE and MAE are the same, then all the errors are of the same size and of the same importance. Comparing RMSE and MAE can be used to determine whether the forecast contains large and infrequent errors.

Table 2 Tableau Linear Forecast Evaluation Summary

Features	**Akaike Information Criterion AIC**	**Mean Absolute Error (tons)**	**Root Mean Squared Error (tons)**	**Forecast Quality**
World Population	**1,640**	**15,146,676**	**33,901,031**	Good
China Final consumption expenditure	2,424	87,024,141,509	142,826,371,542	Ok
India Final consumption expenditure	2,311	30,667,322,232	42,825,338,429	Poor
US Final consumption expenditure	2,445	119,595,841,795	177,635,400,623	Good
CO_2 Emissions FF World (Gt CO_2)	-26	0.57	0.68	Poor
Global Means SurfaceTemp Chg 12 Mth Avg	-181	0.105	0.132	Ok

It can be concluded that the World Population feature is the best statistical fit given the lowest AIC for predicting CO_2 Carbon Emissions future values. The low AIC uses the expected likelihood from traits in the input features. World population growth is fairly predictable, while emissions growth is much more uncertain.

A closer look at the carbon price input data was done in Tableau to farther visualize the linear relationships.

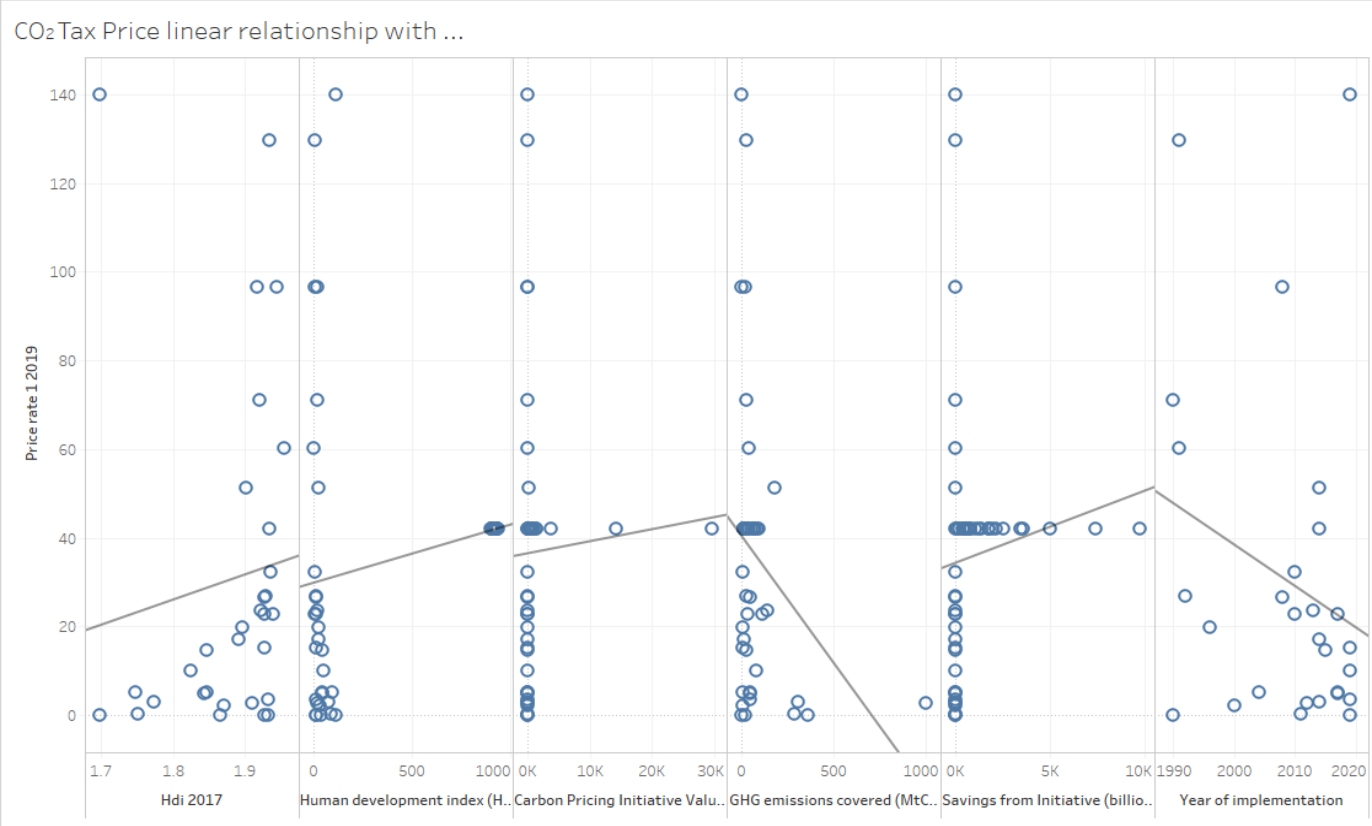

Figure 5 Economic indicators, initiatives value, initiatives savings, implementation year, and CO_2 emissions correlation with price rate (carbon tax)

Figure 5, provides a graphical view of the data relationships between several features and carbon tax prices for initiatives that have been undertaken. These features have a positive linear movement that can be seen against the CO2 tax price: Human Development Indicator (HDI), Carbon Pricing Initiative Value, and Savings from Initiatives data. Appendix A contains feature definitions. Because the graphs show a positive linear relationship between the Carbon Price and these features, they are good starter predictors of the Carbon Tax given the data available. These are the features that are used in the Weka modeling.

Weka

The research leverages several different Tree Based Classifiers as the forecasting models in Weka: DecisionStump, REPTree and RANDTree Classifiers along with ZeroR. They are compared for accuracy and proficiency when assessing the predictions for each dataset. A

decision tree is a supervised machine learning model used to predict a target value after it learns or derives decision rules from input features (Frank et al., 2011). Note, supervised machine learning is the process of an algorithm learning from the training dataset and can be thought of as a teacher supervising the learning process.

- DecisionStump is a machine learning model that builds a one-level decision tree. It makes a prediction based on the value of a single input feature (i.e., US Mining, Manufacturing, Utilities).
- REPTree is a machine learning model that builds a decision or regression tree using information gain and pruning (APPENDIX A)
- RANDTree is a machine learning model that builds a tree that considers a given number of random features at each node
- ZeroR is a machine learning model that is ***rule-based*** and predicts the majority class (if nominal) or the average value (if numeric). It is used to determine a baseline performance as a benchmark for other classifier methods to compare against.

In Weka, a preprocess activity was conducted for attribute selection to improve accuracy and information gain. The original dataset contains many features where some are more relevant than others and using attribute selection preprocessing helps identify the most relevant features needed to produce quality predictions. The two classifiers used are CfsSubsetEval with BestFirst and with RandomSubset. The data attributes (7) selected by the attribute selection models for dataset 1 are: Human development index (HDI), Year of implementation, Year of abolishment, GHG emissions covered [$MtCO_2e$], Carbon Pricing Initiative Value [billion US$], Savings from Initiative [billion US$], Price_rate_1_2019 (class) as defined in Table 13 Data Analysis Terms

This data attributes (8) selected by the attribute selection models for datasets 2-4 are: Country_Final consumption expenditure, Country_Exports of goods and services, Country_Gross Domestic Product (GDP), Country_Agriculture, hunting, forestry, fishing (ISIC A-B), Country_Mining, Manufacturing, Utilities (ISIC C-E), Country_Manufacturing (ISIC D), Global Means_SurfaceTemp_Chg_12_Mth_Avg, and CO_2_Emissions_FF_Country ($GtCO_2$) – (class attribute). These attributes are defined in Table 13 Data Analysis TermsAPPENDIX A.

Note, each method will be set to run with a stratified 10-fold cross-validation which is considered the standard way for predicting the error rate of a learning technique against a single fixed sample dataset (Witten, Ian, Data Mining, 3thE, p 153).

Empirical Analysis

Each study performs Regression tree analysis with the model ZeroR used as the baseline. A Regression tree analysis is when the predicted outcome can be considered a real number (e.g. the price assessed for carbon or property tax, or a car repair's turnaround time).

A decision tree classifier output is similar to a hierarchical tree diagram with the subordinate or lower level nodes representing classification outputs or decisions. The objective for selecting a decision tree is to gain advantage from finding attributes that produce the most efficiently organized tree, sometimes measured via the best information gain (APPENDIX A).

Compare Models of Carbon Tax Predictions

Compare Weka Models

Carbon tax proposals can be evaluated via policy, political, economic, scientific, or analytics perspectives. For example, from a scientific and economic perspective, the IPCC estimates a tax must range from $135 to $5,500 per ton through 2030 and from $690 to $27,000 per ton through 2100 to be effective against the climate change problem. (IPCC special report, 2019). In this work the carbon tax is generated using decision trees learned via Weka. For the ***Better Performing Model***, the model with bigger correlation and smaller error estimates is selected as a candidate for the solution's recommendation. Decision tree learning algorithms try to generate an efficient tree, so smaller trees are preferred over larger trees, other things being equal. The differences in the algorithms include how each choice (branch node) is determined. There could be large information gain, or low cross-validated error, among other possibilities.

In order to evaluate the forecasting effect of each model more objectively, R (correlation coefficient), MAE (mean absolute error), RMSE (root mean square error) and RRSE (root relative squared error) are applied to compare the forecasting accuracy or fit of each model: ZeroR, DecisionStump, REPTree, RANDTree. Note, the RMSE (root mean square error) statistic is the first measure observed for best fit. Also, we observe the difference between RMSE and MAE to see if the forecast contains large but infrequent errors. Larger differences between RMSE and MAE indicate more inconsistency in the error size.

The **Carbon Tax study** used the classifiers found in Table 3. During an analytical review of the Weka run for RANDTree, the Visualize Classifier Error interface was used to get insights on the feature relationships.

Table 3 shows the evaluation of the learning algorithms on Dataset 1. The MAE and RMSE statistics in Table 3, show the goodness of fit for each of the models. From this, the MAE, RMSE and RRSE of the DecisionStump model are the smallest of all the models. It can be concluded that DecisionStump model is the best statistical fit. Given other things being equal, the simpler model (smaller tree) is preferred in addition to having the best results. DecisionStump has the smallest tree, while RANDTree has the largest tree (1 vs. 55 nodes).

Table 3 Dataset 1 Carbon Tax Rate/Price Model Evaluation Summary

Classifier	Correlation Coefficient(r)	Mean Absolute Error ($)	Root Mean Squared Error ($)	Root Relative Squared Error
ZeroR	-0.5166	17.7934	29.085	100%
DecisionStump	**0.4202**	**13.0738**	**25.7317**	**88.4706%**
REPTree	-0.3984	17.5197	29.1267	100.1434%
RANDTree	0.2175	15.8759	35.3578	127.227 %

However there are other factors that influence the final selection for the study's solution. It is important to keep "a subject matter expert in the loop" to ensure that patterns detected in machine learning makes sense and are not just detecting coincidental patterns. Note, the RANDTree model may be best suited to be more effective and practical for CO_2 Tax predictions given the available information for making predictions. Table 4, shows what the decision tree learned for predicting Carbon Tax Rate by each classifier. Weka's display of decision trees is a bit sidewise, with the top of the tree at the left most indentation. The pruned decision tree shown in text, show the class value (carbon tax rate) predicted. Note, pruning is a technique in machine learning and search algorithms to reduce the size of decision trees by removing sections of the tree that provide little power to classify instances. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting. (Frank et al., 2011).

DecisionStump's ruleset uses a predictor: *Carbon Pricing Initiative Value* that has a positive relationship with Carbon Price and provide a simple prediction approach that targets a baseline initiative value and determines which predicted price to assign when the Carbon Pricing Initiative Value is above or below that value amount. A larger value for the carbon pricing initiative results in a larger carbon tax rate. This is perhaps an over simplistic model; given that the desire to raise more funding (revenue) could easily result in a policy decision to charge a higher tax and that could be a very simple non-analytical explanation.

Even though REPTree has a greater error than DecisionStump, the larger tree has the potential to tell us more about what is happening in the data. REPTree, like DecisionStump, starts with the size of the initiative. In reading REPTree models, such as seen in **Table 4**, a leaf node such as: Carbon Pricing Initiative Value [billion US$] >= 0.52: 24.61 (5/3.35) [0/0], 24,61, 5/0, means that 5 instances reached the leaf correctly and 0 is incorrectly classified. Also, the tax rate predicted is $24.61. Below the top node, the REPTree model uses the feature HDI, which is a composite measure of a on country's development based on several factors. This makes sense in that more developed countries may be carrying more economic activity, perhaps leading to more emissions, and can better afford a higher tax. However, the next to last leaf predicts a high carbon tax for countries with low HDI rank (greater than 101.5). It would be interesting to try to explore whether that is some sort of anomaly or if there is a good explanation for that pattern.

If a more detailed ruleset is required, the RANDTree model produced a decision tree that offers a wide range of tax rates base on additional features (HDI_2017) that are considered good predictors. These predictors have a positive linear relationship with Carbon Price. There is also much greater branching based on different values, perhaps creating overfitting. The decision tree can be forwarded to the subject matter experts who can better determine how accurate the

ruleset conditions are as they relate to the predicted Carbon price. Also, it should be added that RANDTree had more bad errors (see RMSE).

Table 4 Dataset 1 Model Output Summary

Classifier	Predictions
ZeroR	Test mode: 10-fold cross-validation, Classifier model (full training set) ZeroR predicts class value: **35.25**142263076923
DecisionStump	Test mode: 10-fold cross-validation, Classifier model (full training set) Carbon Pricing Initiative Value [billion US$] <= 1.1884000000000001: **21.369521103448278** Carbon Pricing Initiative Value [billion US$] > 1.1884000000000001: **46.43406552777779** Carbon Pricing Initiative Value [billion US$] is missing: **35.251422630769234**
REPTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 9 Carbon Pricing Initiative Value [billion US$] < 1.19 \| Human development index (HDI) < 101.5 \| \| Carbon Pricing Initiative Value [billion US$] < 0.52 | | | Human development index (HDI) < 29.5: **18.49** (4/59.67) [6/1325.78] | | | Human development index (HDI) >= 29.5: **5.09** (6/11.36) [3/39.28] | | Carbon Pricing Initiative Value [billion US$] >= 0.52: **24.61** (5/3.35) [0/0] \| Human development index (HDI) >= 101.5: **53.2** (4/2646) [1/196] Carbon Pricing Initiative Value [billion US$] >= 1.19: **46.43** (24/497.69) [12/73.84]
RANDTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 55 Carbon Pricing Initiative Value [billion US$] < 1.19 \| Human development index (HDI) < 101.5 \| \| Human development index (HDI) < 28 \| \| \| Carbon Pricing Initiative Value [billion US$] < 0.03: **96.68** (1/0) | | | Carbon Pricing Initiative Value [billion US$] >= 0.03 \| \| \| \| Carbon Pricing Initiative Value [billion US$] < 0.8 | | | | | Human development index (HDI) < 18 | | | | | | HDI_2017 < 1.94 | | | | | | | Human development index (HDI) < 8.5: **32.19** (1/0) | | | | | | | Human development index (HDI) >= 8.5: **26.94** (1/0) | | | | | | HDI_2017 >= 1.94: **22.94** (1/0) | | | | | Human development index (HDI) >= 18 | | | | | | Human development index (HDI) < 25.5: **19.84** (1/0) | | | | | | Human development index (HDI) >= 25.5: **17.21** (1/0) | | | | Carbon Pricing Initiative Value [billion US$] >= 0.8 \| \| \| \| \| Carbon Pricing Initiative Value [billion US$] < 1.04

	| | | | | | HDI_2017 < 1.93: **5.09** (3/51.84) | | | | | | HDI_2017 >= 1.93 | | | | | | | Human development index (HDI) < 9.5: **3.7** (1/0) | | | | | | | Human development index (HDI) >= 9.5: **0** (1/0) | | | | | Carbon Pricing Initiative Value [billion US$] >= 1.04 \| \| \| \| \| \| Carbon Pricing Initiative Value [billion US$] < 1.09: **22.91** (1/0) | | | | | | Carbon Pricing Initiative Value [billion US$] >= 1.09 \| \| \| \| \| \| \| HDI_2017 < 1.92: **23.51** (1/0) \| \| \| \| \| \| \| HDI_2017 >= 1.92: **26.73** (1/0) \| \| Human development index (HDI) >= 28 \| \| \| Carbon Pricing Initiative Value [billion US$] < 0.01 | | | | HDI_2017 < 1.87: **0.22** (2/0.02) | | | | HDI_2017 >= 1.87: **2.29** (1/0) | | | Carbon Pricing Initiative Value [billion US$] >= 0.01 \| \| \| \| Human development index (HDI) < 60.5 \| \| \| \| \| Carbon Pricing Initiative Value [billion US$] < 0.16: **5.08** (2/0.01) | | | | | Carbon Pricing Initiative Value [billion US$] >= 0.16 \| \| \| \| \| \| HDI_2017 < 1.84: **10** (1/0) \| \| \| \| \| \| HDI_2017 >= 1.84: **14.61** (1/0) \| \| \| \| Human development index (HDI) >= 60.5 \| \| \| \| \| Human development index (HDI) < 82: **3.01** (1/0) \| \| \| \| \| Human development index (HDI) >= 82: **5.27** (1/0) \| Human development index (HDI) >= 101.5 \| \| Human development index (HDI) < 510.5: **70** (2/4900) \| \| Human development index (HDI) >= 510.5: **42** (3/0) Carbon Pricing Initiative Value [billion US$] >= 1.19 | Human development index (HDI) < 17 | | Human development index (HDI) < 1.5: **60.27** (1/0) | | Human development index (HDI) >= 1.5 | | | Carbon Pricing Initiative Value [billion US$] < 2.22 \| \| \| \| Human development index (HDI) < 8.5: **96.68** (1/0) \| \| \| \| Human development index (HDI) >= 8.5: **71.12** (1/0) \| \| \| Carbon Pricing Initiative Value [billion US$] >= 2.22: **129.74** (1/0) | Human development index (HDI) >= 17 | | Carbon Pricing Initiative Value [billion US$] < 6.02: **2.65** (1/0) \| \| Carbon Pricing Initiative Value [billion US$] >= 6.02 | | | Human development index (HDI) < 462.5: **51.16** (1/0) | | | Human development index (HDI) >= 462.5: **42** (30/0)

Compare Models of Carbon (CO_2) Emissions Predictions

The input features used for Dataset 2-4 by the models are: Country_Final consumption expenditure, Country_Exports of goods and services, Country_Gross Domestic Product (GDP), Country_Agriculture, hunting, forestry, fishing (ISIC A-B), Country_Mining, Manufacturing, Utilities (ISIC C-E), Country_Manufacturing (ISIC D), Global Means_SurfaceTemp_Chg_12_Mth_Avg, and CO_2_Emissions_FF_Country ($GtCO_2$) – (class attribute).

The **Carbon (CO_2) Emissions: United States (US) study** used the same classifiers as used for carbon price prediction. Table 5 shows the evaluation of the learning algorithms on Dataset 2. The MAE, RMSE and RRSE of the REPTree model are the smallest of all the models. While the MAE and RMSE statistics in Table 5 Dataset 2 US Carbon (CO2) Emissions Model Evaluation Summary are very close for REPTree and RANDTree, the difference between these on RMSE suggests that REPTree has fewer unusually large errors. The goodness of fit is ranked as follows: ***REPTree*** > *RANDTree* > *DecisionStump* >*ZeroR*. REPTree model statistically has the better forecasting performance than the RANDTree model. Furthermore, the RANDTree model's tree size is 77 to 21 for REPTree. Even though RANDree has a greater error than REPTree, the larger tree has the potential to tell more about what is happening in the data.

Table 5 Dataset 2 US Carbon (CO_2) Emissions Model Evaluation Summary

Classifier	Correlation Coefficient(r)	Mean Absolute Error (Gt CO_2)	Root Mean Absolute Error (Gt CO_2)	Root Relative Squared Error
ZeroR	0.5055	0.5245	0.5979	100%
DecisionStump	0.8612	0.2585	0.2959	49.4864%
REPTree	**0.9244**	**0.1836**	**0.2232**	**37.3256%**
RANDTree	0.9108	0.1843	0.247	41.3159%

Table 6 **Dataset 2 Model Output Summary** show the predicted CO_2 Emissions derived by each classifier. The pruned decision tree shown in text, shows the CO_2 Emissions value predicted at the leaf nodes. It can be concluded that the REPTree model is the best statistical fit. Given things being the same, the simpler model (smaller tree) is preferred in addition to having the best results. It is important to keep "a subject matter expert in the loop" to ensure that patterns detected in machine learning makes sense and are not just detecting coincidental patterns.

Both DecisionStump and RANDTree start with US Gross Domestic Product (GDP) using the US dataset and both show greater error than REPTree. The US REPTree model chose US_Agriculture, hunting, forestry, fishing as its root node. Both of these meet the "eye test" of making sense. The US is the largest economy in the world based on GDP and the second largest exporter in the world which fuels global consumption. The tree nodes show that predictions are based on the US market value of goods and services (GDP), production and agriculture, trade (export and imports), and consumption expenditures by its citizens and businesses.

In addition to the root node, the higher a feature is in the tree, the more it is being used. The more often the feature shows up in the tree, the more it is being used. REPTree and RANDTree favor the Final Consumption Expenditures feature used in many nodes in the respective trees. The Final Consumption Expenditures feature is used to show the expenditures incurred by household units on goods and services. This makes sense since the expenditures represent the demand side of the economy where people consume products manufactured and services that emit CO_2 pollution. This is an interesting pattern since it is supported by the supply and demand model.

Interestingly, the RANDTree, Global Surface temperature occurs in many places in the tree; it appears to create a reverse affect because consumption would impact CO_2 emission which than impacts Surface Temperature. This anomaly appears to show a reverse in cause and effect where the machine learning detected a pattern that was the reverse of the actual causality.

Table 6 Dataset 2 Model Output Summary

Classifier	Predictions
ZeroR	Test mode: 10-fold cross-validation, Classifier model (full training set) ZeroR predicts class value: **5.204583333333334**
DecisionStump	Test mode: 10-fold cross-validation, Classifier model (full training set) US_Gross Domestic Product (GDP) <= 6.339228E12: **4.645** US_Gross Domestic Product (GDP) > 6.339228E12: **5.678076923076923** US_Gross Domestic Product (GDP) is missing: **5.204583333333335**
REPTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 21 US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 82452461538.5 | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.14: 4.45 (4/0.02) [3/0.01] | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.14 | | US_Final consumption expenditure < 2540657500000 | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.21: **4.89** (2/0) [0/0] | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.21: **4.69** (2/0) [2/0.02] | | US_Final consumption expenditure >= 2540657500000: **4.56** (4/0.04) [2/0.01] US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 82452461538.5 | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.44: **5.2** (3/0.01) [1/0.11] | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.44 | | US_Exports of goods and services < 1974637500000 | | | US_Final consumption expenditure < 12089031500000 | | | | US_Final consumption expenditure < 7915549000000: **5.51** (3/0.01) [4/0.23] | | | | US_Final consumption expenditure >= 7915549000000 | | | | | US_Final consumption expenditure < 9767722000000: **5.96** (3/0) [1/0.01] | | | | | US_Final consumption expenditure >= 9767722000000: **6.1** (3/0) [1/0] | | | US_Final consumption expenditure >= 12089031500000: **5.71** (3/0.03) [0/0] | | US_Exports of goods and services >= 1974637500000 | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 180250000000: **5.32** (2/0) [1/0] | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 180250000000: **5.52** (3/0.01) [1/0]
RANDTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 77 US_Gross Domestic Product (GDP) < 6339228000000 | US_Exports of goods and services < 404272000000 | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 29128654871.5: **4.34** (2/0) | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 29128654871.5 | | | US_Gross Domestic Product (GDP) < 3032174500000 | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 45920953077 | | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 45632142051.5 | | | | | | US_Final consumption expenditure < 1538659000000 | | | | | | | US_Final consumption expenditure < 1094508500000: **4.56** (1/0) | | | | | | | US_Final consumption expenditure >= 1094508500000: **4.6** (2/0)

	| | | | | | US_Final consumption expenditure >= 1538659000000: **4.74** (1/0) | | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 45632142051.5: **4.4** (1/0) | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 45920953077 | | | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.21: **4.89** (2/0) | | | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.21 | | | | | | US_Final consumption expenditure < 1646508500000: **4.77** (1/0) | | | | | | US_Final consumption expenditure >= 1646508500000: **4.72** (1/0) | | | US_Gross Domestic Product (GDP) >= 3032174500000 | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 66797864359 | | | | | US_Final consumption expenditure < 2751392000000: **4.3** (1/0) | | | | | US_Final consumption expenditure >= 2751392000000: **4.34** (1/0) | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 66797864359 | | | | | US_Manufacturing (ISIC D) < 862700916931.5 | | | | | | US_Final consumption expenditure < 2784030000000: **4.53** (1/0) | | | | | | US_Final consumption expenditure >= 2784030000000: **4.48** (3/0) | | | | | US_Manufacturing (ISIC D) >= 862700916931.5: **4.68** (1/0) | US_Exports of goods and services >= 404272000000 | | US_Exports of goods and services < 528081000000 | | | US_Final consumption expenditure < 4304327000000: **4.89** (1/0) | | | US_Final consumption expenditure >= 4304327000000: **4.95** (1/0) | | US_Exports of goods and services >= 528081000000 | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 88216615384.5: **5.07** (1/0) | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 88216615384.5: **5.12** (1/0) US_Gross Domestic Product (GDP) >= 6339228000000 | US_Final consumption expenditure < 12924420500000 | | US_Final consumption expenditure < 6260909000000 | | | US_Final consumption expenditure < 5677045000000 | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 91892307692: **5.29** (1/0) | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 91892307692: **5.18** (1/0) | | | US_Final consumption expenditure >= 5677045000000 | | | | US_Final consumption expenditure < 5966700000000: **5.39** (1/0) | | | | US_Final consumption expenditure >= 5966700000000: **5.45** (1/0) | | US_Final consumption expenditure >= 6260909000000 | | | US_Exports of goods and services < 973290500000 | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 108767076923 | | | | | US_Final consumption expenditure < 6951086500000: **5.71** (1/0) | | | | | US_Final consumption expenditure >= 6951086500000: **5.75** (1/0) | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 108767076923: **5.64** (1/0) | | | US_Exports of goods and services >= 973290500000 | | | | US_Exports of goods and services < 1527304500000 | | | | | US_Final consumption expenditure < 9767722000000 | | | | | | US_Exports of goods and services < 1030406500000 | | | | | | | US_Final consumption expenditure < 8116993000000: **5.83** (1/0) | | | | | | | US_Final consumption expenditure >= 8116993000000 | | | | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 97700000000: **5.94** (1/0)

	| | | | | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 97700000000: **5.9** (1/0) | | | | | | US_Exports of goods and services >= 1030406500000: **6** (2/0) | | | | | US_Final consumption expenditure >= 9767722000000 | | | | | | US_Final consumption expenditure < 11021754000000: **6.12** (2/0) | | | | | | US_Final consumption expenditure >= 11021754000000: **6.05** (1/0) | | | | US_Exports of goods and services >= 1527304500000 | | | | | US_Final consumption expenditure < 12089031500000: **6.13** (1/0) | | | | | US_Final consumption expenditure >= 12089031500000 | | | | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.71: **5.93** (1/0) | | | | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.71 | | | | | | | US_Exports of goods and services < 1714138000000: **5.5** (1/0) | | | | | | | US_Exports of goods and services >= 1714138000000: **5.7** (1/0) | US_Final consumption expenditure >= 12924420500000 | | US_Final consumption expenditure < 15166808500000 | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.77: **5.36** (1/0) | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.77 | | | | US_Final consumption expenditure < 14648234000000 | | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) < 208200000000: **5.57** (2/0) | | | | | US_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 208200000000: **5.51** (1/0) | | | | US_Final consumption expenditure >= 14648234000000: **5.41** (1/0) | | US_Final consumption expenditure >= 15166808500000: **5.29** (2/0)

The **Carbon (CO2) Emissions: China study** used the classifiers found in Table 7. Table 7 shows the evaluation of the learning algorithms on Dataset 3. The R, MAE, RMSE and RRSE of the RANDTree model is the smallest of all the models. The goodness of fit is ranked as follows: RANDTree > REPTree > DecisionStump >ZeroR. Furthermore, the RANDTree model's tree size is 47 to 13 for REPTree. Even though REPTree has a greater error than RANDTree, all thing being the samel, it would be preferred because it is the smaller tree.

Table 7 Dataset 3 China Economic Indicators, GHG emissions, Land Temperature and Population Model Evaluation Summary

Classifier	Correlation Coefficient(r)	Mean Absolute Error (Gt CO_2)	Root Mean Absolute Error (Gt CO_2)	Root Relative Squared Error
ZeroR	-0.5207	2.7166	3.2472	100%
DecisionStump	0.9472	0.8501	1.0178	31.3453%
REPTree	0.9604	0.5651	0.8879	27.3426%
RANDTree	**0.994**	**0.2599**	**0.3506**	**10.7963%**

Table 8 Dataset 3 Model Output Summary shows the CO_2 Emissions derived by each classifier. The pruned decision tree shown in text, shows the CO_2 Emissions value predicted. It is important to keep "a subject matter expert in the loop" to ensure that patterns detected in machine learning makes sense and are not just detecting coincidental patterns. But more importantly the decision tree is large and contains more complexity than the smaller trees.

The root nodes selected for the China decision tree models are Manufacturing in DecisionStump, Final consumption expenditures in REPTree and Exports of goods and services in RANDTree. Note, the higher a feature is in the tree for more it is used. Also, when a feature appears often throughout the tree, the more it is used. RANDTree is a larger tree with added complexity. RANDTree favors the Exports of Goods and Services feature, using it in many nodes. Exports of Goods and Services is used to show sales, barter, gifts or grants, of goods and services from residents within a country to non-residents outside the country. This makes sense since the exports of goods and services represent the supply side of China's economy where its population produces more products and services that are used in other countries and emit CO_2 pollution. This is an interesting pattern since it is also supported by the supply and demand model.

China is the second largest economy in the world based on GDP and the largest exporter in the world. The tree nodes show that the predictions are based on the China market value of goods and services (GDP), production, trade (export and imports), and consumption expenditures by its citizens and businesses. Additionally, note that the trees for China does not use Agriculture for predicting, where the US and India do.

Table 8 Dataset 3 Model Output Summary

Classifier	Predictions
ZeroR	Test mode: 10-fold cross-validation, Classifier model (full training set) ZeroR predicts class value: **4.082708333333334**
DecisionStump	Test mode: 10-fold cross-validation, Classifier model (full training set) China_Manufacturing (ISIC D) <= 1.76845147601E12: **6.701666666666667** China_Manufacturing (ISIC D) > 1.76845147601E12: **9.978750000000003** China_Manufacturing (ISIC D) is missing: **2.233235294117647**
REPTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 13 China_Final consumption expenditure < 1157179910903 | China_Exports of goods and services < 62903764474.5 | | China_Exports of goods and services < 22983178733.5: **1.15** (7/0.07) [4/0.04] | | China_Exports of goods and services >= 22983178733.5 | | | China_Exports of goods and services < 28724647636: **1.63** (2/0) [2/0.05] | | | China_Exports of goods and services >= 28724647636: **2.29** (3/0.02) [4/0.07] | China_Exports of goods and services >= 62903764474.5 | | China_Final consumption expenditure < 927759170906: **3.3** (8/0.12) [3/0.02] | | China_Final consumption expenditure >= 927759170906: **4.89** (2/0.12) [0/0] China_Final consumption expenditure >= 1157179910903 | China_Final consumption expenditure < 2602716874233: **7** (3/0.48) [2/0.73] | China_Final consumption expenditure >= 2602716874233: **9.98** (7/0.28) [1/0.04]
RANDTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 47 China_Exports of goods and services < 690224464976 | China_Final consumption expenditure < 293630940445 | | China_Exports of goods and services < 26697637304 | | | China_Gross Domestic Product (GDP) < 169183704714 | | | | China_Final consumption expenditure < 96714277266.5: **0.91** (5/0.01) | | | | China_Final consumption expenditure >= 96714277266.5: **1.17** (2/0) | | | China_Gross Domestic Product (GDP) >= 169183704714 | | | | China_Exports of goods and services < 10259702055: **1.31** (1/0) | | | | China_Exports of goods and services >= 10259702055: **1.52** (6/0.01) | | China_Exports of goods and services >= 26697637304 | | | China_Mining, Manufacturing, Utilities (ISIC C-E) < 122765612697.5 | | | | China_Final consumption expenditure < 203870738239.5: **2.01** (2/0) | | | | China_Final consumption expenditure >= 203870738239.5: **1.81** (1/0) | | | China_Mining, Manufacturing, Utilities (ISIC C-E) >= 122765612697.5 | | | | China_Exports of goods and services < 42361849694: **2.21** (1/0) | | | | China_Exports of goods and services >= 42361849694: **2.45** (4/0.01) | China_Final consumption expenditure >= 293630940445 | | China_Final consumption expenditure < 927759170906 | | | China_Final consumption expenditure < 397182588515 | | | | China_Final consumption expenditure < 311126533747.5: **2.69** (1/0) | | | | China_Final consumption expenditure >= 311126533747.5: **2.97** (2/0.01) | | | China_Final consumption expenditure >= 397182588515 | | | | China_Exports of goods and services < 332403067496: **3.4** (7/0) | | | | China_Exports of goods and services >= 332403067496: **3.85** (1/0)

	| | China_Final consumption expenditure >= 927759170906 | | | China_Exports of goods and services < 546192161786: **4.54** (1/0) | | | China_Exports of goods and services >= 546192161786: **5.23** (1/0) China_Exports of goods and services >= 690224464976 | China_Final consumption expenditure < 2733866842728.5 | | China_Gross Domestic Product (GDP) < 3152142511185 | | | China_Exports of goods and services < 882309834696: **5.89** (1/0) | | | China_Exports of goods and services >= 882309834696: **6.52** (1/0) | | China_Gross Domestic Product (GDP) >= 3152142511185 | | | China_Gross Domestic Product (GDP) < 4075200773700.5: **7.03** (1/0) | | | China_Gross Domestic Product (GDP) >= 4075200773700.5 | | | | China_Agriculture, hunting, forestry, fishing (ISIC A-B) < 500290293022.5: **7.55** (1/0) | | | | China_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 500290293022.5: **7.99** (1/0) | China_Final consumption expenditure >= 2733866842728.5 | | China_Gross Domestic Product (GDP) < 6836623974265.5: **8.77** (1/0) | | China_Gross Domestic Product (GDP) >= 6836623974265.5 | | | China_Agriculture, hunting, forestry, fishing (ISIC A-B) < 874591637925.5 | | | | China_Exports of goods and services < 2090689083893.5: **9.73** (1/0) | | | | China_Exports of goods and services >= 2090689083893.5: **10.02** (1/0) | | | China_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 874591637925.5 | | | | China_Mining, Manufacturing, Utilities (ISIC C-E) < 3974668011032.5: **10.21** (4/0) | | | | China_Mining, Manufacturing, Utilities (ISIC C-E) >= 3974668011032.5: **10.48** (1/0)

The **Carbon (CO_2) Emissions: India study** used the classifiers found in **Table 9** shows the evaluation of the learning algorithms on Dataset 4. From **Table 9,** the MAE, RMSE and RRSE of the RANDTree model are the smallest of all the models. The goodness of fit is ranked as follows: ***RANDTree*** > *REPTree* > *DecisionStump* >*ZeroR.* However, the RANDTree model's tree size is 51 to 21 for REPTree. All things being equal, a smaller tree is preferred, but the differences in errors between REPTree and RANDTree are rather significant here.

Table 9 Dataset 4 India Economic Indicators, GHG emissions, Land Temperature and Population Model Evaluation Summary.

Table 9 shows the evaluation of the learning algorithms on Dataset 4. From **Table 9**, the MAE, RMSE and RRSE of the RANDTree model are the smallest of all the models. The goodness of fit is ranked as follows: ***RANDTree*** > *REPTree* > *DecisionStump* >*ZeroR.* However, the RANDTree model's tree size is 51 to 21 for REPTree. All things being equal, a

smaller tree is preferred, but the differences in errors between REPTree and RANDTree are rather significant here.

Table 9 Dataset 4 India Economic Indicators, GHG emissions, Land Temperature and Population Model Evaluation Summary

Classifier	Correlation Coefficient(r)	Mean Absolute Error (Gt CO_2)	Root Mean Absolute Error (Gt CO_2)	Root Relative Squared Error
ZeroR	-0.5106	0.5609	0.6869	100%
DecisionStump	0.8512	0.3077	0.3505	51.0229%
REPTree	0.9663	0.1212	0.1724	25.105%
RANDTree	**0.9876**	**0.0793**	**0.1051**	**15.2951%**

Table 10 Dataset 3 Model Output Summary shows the CO_2 Emissions derived by each classifier. The pruned decision tree shown in text, show the class value predicted. As usual, it is important to keep "a subject matter expert in the loop" to ensure that patterns detected in machine learning makes sense and are not just detecting coincidental patterns.

The root nodes selected for the India decision tree models are Agriculture, Hunting, Forestry, and Fishing in DecisionStump, Manufacturing in REPTree and Exports of goods and services in RANDTree. The REPTree tree relies heavily on Exports, and Final Consumption, with some use of Mining, Manufacturing and Utilities. The RANDTree nodes consists of those used in REPTRee plus Gross Domestic Product (GDP), with some use of Agriculture, Hunting, Forestry, and Fishing,and Mean Surface Temperature (12 mos.). . Note, the higher a feature is in the tree for more it is used. Also, when a feature appears often throughout the tree, the more it is used.

India is the 5th largest economy in the world based on GDP and the 18th largest exporter in the world. The tree nodes show that the predictions are based on the India market value of

goods and services (GDP), production and agriculture, trade (export and imports), and consumption expenditures by its citizens and businesses.

Exports of Goods and Services measures the amount of products or services produced or manufactured by a given country that are sent to non-residents outside the country. This makes sense since the exports of goods and services represent the supply side of India's economy. Its huge population produces more products and services that are used in other countries and emits CO_2 pollution. India and China are similar since they are supported by the supply and demand model.

Table 10 Dataset 3 Model Output Summary

Classifier	**Predictions**
ZeroR	Test mode: 10-fold cross-validation, Classifier model (full training set) ZeroR predicts class value: **0.9245833333333334**
DecisionStump	Test mode: 10-fold cross-validation, Classifier model (full training set) India_Agriculture, hunting, forestry, fishing (ISIC A-B) <= 1.514869077675E11: **0.5913888888888889** India_Agriculture, hunting, forestry, fishing (ISIC A-B) > 1.514869077675E11: **1.9241666666666672** India_Agriculture, hunting, forestry, fishing (ISIC A-B) is missing: **0.9245833333333334**
REPTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 21 India_Manufacturing (ISIC D) < 165677126288 | India_Exports of goods and services < 22304889181 | | India_Final consumption expenditure < 167104997621 | | | India_Final consumption expenditure < 72920003064.5: **0.21** (3/0) [0/0] | | | India_Final consumption expenditure >= 72920003064.5: **0.28** (4/0) [6/0] | | India_Final consumption expenditure >= 167104997621 | | | India_Final consumption expenditure < 191813105293: **0.4** (2/0) [1/0] | | | India_Final consumption expenditure >= 191813105293: **0.52** (3/0) [1/0.01] | India_Exports of goods and services >= 22304889181 | | India_Manufacturing (ISIC D) < 73024773190.5 | | | India_Exports of goods and services < 36633449386: **0.69** (3/0) [2/0.01] | | | India_Exports of goods and services >= 36633449386: **0.91** (3/0) [2/0.01] | | India_Manufacturing (ISIC D) >= 73024773190.5 | | | India_Final consumption expenditure < 449742078415.5: **1.06** (3/0) [1/0] | | | India_Final consumption expenditure >= 449742078415.5: **1.22** (2/0) [1/0.01] India_Manufacturing (ISIC D) >= 165677126288 | India_Final consumption expenditure < 1173482376924.5: **1.61** (3/0.02) [1/0.03] | India_Final consumption expenditure >= 1173482376924.5

	| | India_Mining, Manufacturing, Utilities (ISIC C-E) < 401309487388: **1.96** (3/0.01) [0/0] | | India_Mining, Manufacturing, Utilities (ISIC C-E) >= 401309487388: **2.37** (3/0.01) [1/0.01]
RANDTree	Test mode: 10-fold cross-validation, Classifier model (full training set) Size of the tree: 51 India_Exports of goods and services < 180595580703 | India_Mining, Manufacturing, Utilities (ISIC C-E) < 73271402366 | | India_Exports of goods and services < 14802798305.5 | | | India_Exports of goods and services < 11668987766 | | | | India_Exports of goods and services < 6300859557: **0.22** (6/0) | | | | India_Exports of goods and services >= 6300859557: **0.29** (5/0) | | | India_Exports of goods and services >= 11668987766 | | | | India_Final consumption expenditure < 181813224875: **0.36** (4/0) | | | | India_Final consumption expenditure >= 181813224875: **0.45** (2/0) | | India_Exports of goods and services >= 14802798305.5 | | | India_Final consumption expenditure < 219600237641.5 | | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg < 0.39: **0.71** (2/0) | | | | Global Means_SurfaceTemp_Chg_12_Mth_Avg >= 0.39: **0.66** (1/0) | | | India_Final consumption expenditure >= 219600237641.5 | | | | India_Manufacturing (ISIC D) < 50930535889: **0.51** (2/0) | | | | India_Manufacturing (ISIC D) >= 50930535889: **0.6** (2/0) | India_Mining, Manufacturing, Utilities (ISIC C-E) >= 73271402366 | | India_Exports of goods and services < 50107240765 | | | India_Gross Domestic Product (GDP) < 372669245741.5 | | | | India_Exports of goods and services < 36350999255: **0.76** (1/0) | | | | India_Exports of goods and services >= 36350999255: **0.81** (1/0) | | | India_Gross Domestic Product (GDP) >= 372669245741.5 | | | | India_Gross Domestic Product (GDP) < 398875769305: **0.88** (1/0) | | | | India_Gross Domestic Product (GDP) >= 398875769305: **0.93** (2/0) | | India_Exports of goods and services >= 50107240765 | | | India_Final consumption expenditure < 449742078415.5 | | | | India_Final consumption expenditure < 390733768960.5 | | | | | India_Gross Domestic Product (GDP) < 452644844985: **0.99** (1/0) | | | | | India_Gross Domestic Product (GDP) >= 452644844985: **1.04** (3/0) | | | | India_Final consumption expenditure >= 390733768960.5: **1.1** (1/0) | | | India_Final consumption expenditure >= 449742078415.5 | | | | India_Agriculture, hunting, forestry, fishing (ISIC A-B) < 135191142420: **1.15** (1/0) | | | | India_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 135191142420: **1.22** (1/0) India_Exports of goods and services >= 180595580703 | India_Exports of goods and services < 399675411760 | | India_Exports of goods and services < 257366458970.5 | | | India_Gross Domestic Product (GDP) < 1061895452921.5: **1.3** (1/0) | | | India_Gross Domestic Product (GDP) >= 1061895452921.5: **1.41** (1/0) | | India_Exports of goods and services >= 257366458970.5 | | | India_Final consumption expenditure < 869328100461.5: **1.57** (1/0) | | | India_Final consumption expenditure >= 869328100461.5: **1.73** (2/0) | India_Exports of goods and services >= 399675411760 | | India_Mining, Manufacturing, Utilities (ISIC C-E) < 401309487388 | | | India_Mining, Manufacturing, Utilities (ISIC C-E) < 393084006832: **2.02** (2/0) | | | India_Mining, Manufacturing, Utilities (ISIC C-E) >= 393084006832: **1.84** (1/0) | | India_Mining, Manufacturing, Utilities (ISIC C-E) >= 401309487388

	| | | India_Agriculture, hunting, forestry, fishing (ISIC A-B) < 358280131744 | | | | India_Exports of goods and services < 447280974078.5: **2.32** (1/0) | | | | India_Exports of goods and services >= 447280974078.5: **2.24** (1/0) | | | India_Agriculture, hunting, forestry, fishing (ISIC A-B) >= 358280131744: **2.45** (2/0)

Overall Comparison

Table 11 shows the evaluation of the best learning algorithms taken from each study. The goodness of fit is ranked as follows: China RANDTree > India RANDTree > US-DecisionStump, based on Root Relative Squared Error, which controls for the difficulty of different datasets.

Table 11 Comparison of CO_2 Emissions Best-Fit Models

Model	Correlation Coefficient(r)	Mean Absolute Error (Gt CO_2)	Root Mean Absolute Error (Gt CO_2)	Root Relative Squared Error
US- Decision Stump	0.8612	0.2585	0.2959	49.4864%
China - RANDTree	**0.994**	0.2599	0.3506	**10.7963%**
India - RANDTree	0.9876	**0.0793**	**0.1051**	15.2951%

It is worth noting, looking back at the models that RANDTree selected Exports of **G**oods and **S**ervices for both India and China align with their high world ranking for Exports of **G**oods and **S**ervices and their large populations used to produce goods and services.

From the analysis conducted in the various studies using the Weka platform and machine learning capabilities, the results from the analysis suggest the machine learning methods used adequately created carbon tax rate predictions and CO_2 emissions predictions for the United States, China, and India.

Model Selection and Conclusion

Carbon taxes from input sources were analyzed and machine learning models were constructed to predict carbon tax and CO_2 emissions. The models (classifiers) performance summary outputs were analyzed for accuracy and statistical fit. The results were used to assess whether the predictions address the problem.

As part of the analysis, a close look was given to the RANDTree model results and it was suggested that it is the most feasible predictive solution for the carbon tax and CO_2 emissions. However, the prediction results for CO_2 emissions by country shows REPTree and DecisonStump could be viable options given the appropriate circumstances. There was a notable and interesting observation discovered when reviewing the CO_2 emissions solutions' predictions and that was the high usage of the Final Consumptions Expenditure feature as decision points when determining predictions. This corresponds with expert beliefs, like David Satterthwaite from International Institute of Environment and Development in the UK. He stated, "Changes in our consumption are the key drivers of global warming more so than increasing the number of people on the planet (Satterthwaite, 2009). Higher consumption is what drives anthropogenic climate change, or the production of greenhouse gases emitted by human activity" (Satterthwaite, 2009). The validity in Satterthwaite's statement pertaining to consumption can be seem in the tree created by the machine learning exercise.

Because the graphs show a strong positive linear relationship between the World CO_2 Emissions and Final Consumptions Expenditures for each country, it can be concluded that as the expenditures of the governments and households maintain an upward momentum or increase the CO_2 Emissions will increase and add more CO_2 tonnage into the atmosphere.

Given the research articles referenced in the Literature Review, this study started with insights that were based on adapting to new circumstances and incorporating lessons learned. There are new ideas and refinements that will lead to better forecasts and predictions through new algorithms and modeling techniques in machine learning. Machine learning will continue to be fine-tuned, adapting to new circumstances and incorporating lessons learned. Existing carbon pricing initiatives are evolving based on past experiences and upcoming initiatives try to learn from these experiences for their design.

The studies and analysis conducted have shown machine learning in Weka can produce relevant Carbon tax predictions and produced favorable results using DecisionStump and RANDTree algorithms. Additionally, the studies and analysis conducted have shown machine learning in Weka and linear regression forecasting in Tableau can adequately produce CO_2 predictions and forecasting projections based on economic, population, and surface temperature data features.

DISCUSSION

Lessons Learned and suggestions for continuing and/or expanding areas research for this study/problem

Many business and social problems are solved using machine learning. The various models produced in this study are comparable to those in other studies and if similar models are created by appropriate government agencies or industrial organizations, it could be the beginning of new possibilities and frontiers for climate change analytics. Once a baseline standard for a global carbon tax is established and accepted globally, the carbon tax models can be used against annual data as it is created and projected.

The International Panel of Climate Change (IPPC) acknowledge the need to response to the greenhouse effect and climate change with an effective cost approach that establishes a tax for CO_2 emissions into atmosphere. People are working and exploring various way to derive a cost for CO_2 pollution using machine learning whether it be through single models or hybrid methodologies.

Another observation to be noted from the study's results, is that the Human Development Indicator data was instrumental in setting conditions for price predictions and will serve as a good predictor for future models. The HDI help establish a tree quantified a fair carbon tax / price across all countries given their degree of development.

From the predictions and forecasting results in this study, it can be suggested that CO_2 emissions are expected to increase if no reduction efforts are put into effect. The impact of that could be predicted. If CO_2 emissions and temperature increase, it can be expected that the planet's polar caps will melt and increase water level that impact shoreline. Machine learning predictions can assist with anticipating where water levels will increase if surface temperature increases.

Additionally, as more water enters the waterway, machine learning and artificial intelligence may be used to continue studies about where and when the additional water will create adverse effects to travel over the waterways. Also, the temperature changes will increase extreme weather conditions that are also beyond human control but present patterns that can help anticipate threating condition before they materialize.

Weka has a good toolset of classifiers, clustering and association capabilities for machine learning. Weka's visualization function is not as developed as what is found in Tableau. Tableau offers better visual analysis where Weka provides data analysis.

The primary lesson learned from this capstone article is that machine learning, linear regression forecasting and data visualization tools were more adequate to produce the data analysis information for this study. These tools complement each other and used the same input features to produce the predictions and forecasts.

Using machine learning tools appear to be a viable method for creating carbon tax models that can be used efficiently and effectively on a global scale. The researched evidence used in this study comes from reliable sources. The evidence is connected in new and innovative ways that expand on how machine learning methods and models help add clarify the climate change story and how carbon tax/price and CO_2 emissions can be forecasted to better support the climate mitigation efforts. There is significant room for improvement and refinement on the modeling method and techniques. Progressively, each passing year allows for improvements and adjustments from lessons learned and innovations.

Lessons learned from this study are:

- The study has shown that carbon tax and CO_2 emissions can proficiently be predicted and forecasted using an open source machine learning platform combined with a commercial visualization tool like Tableau.
- The analysis shown via Tableau provided insights that show highly correlated input data for CO_2 emissions and identified a good set of predictors for CO_2 emissions.
- On average, the analysis also determined that the DecisionStump classifier is better classifier, however, the RANDTree classifier produced a deeper tree that show the complexity of the Carbon Tax problem. It can be concluded from the empirical analysis that the economic features are good predictors of Carbon Price and CO_2

Emissions. The open data made available through World Bank.org was instrumental to this study and can be used for future work similar to this study.

On the basis of the evidence presented; the machine learning and forecasting capabilities can be combined to effectively produce carbon tax rate and CO2 emissions prediction and forecasting to help Climate change mitigation.

BIBLIOGRAPHY

Table 12: References

1 Carbon Tax Basics. (2018, February 16). Retrieved July 3, 2019, from https://www.c2es.org/content/carbon-tax-basics/

2 Chan, K., & Tsay, R. S. (2012). Discussion of "feature matching in time series modeling" by Y. xia and H. tong. *Statistical Science, 26*(1), 53-56. doi:10.1214/11-STS345B

3 Cote, N. (2019, January 17). To curb climate change, we have to suck carbon from the sky. But how? Retrieved July 3, 2019, from https://www.nationalgeographic.com/environment/2019/01/carbon-capture-trees-atmosphere-climate-change/

4 Dai, S., Niu, D., & Han, Y. (2018). Forecasting of energy-related CO_2 emissions in China based on GM(1,1) and least squares support vector machine optimized by modified shuffled frog leaping algorithm for sustainability. *Sustainability, 10*(4), 958. doi:10.3390/su10040958

5 Drawdown: 100 Solutions to Reverse Global Warming. (2019, June 12). Retrieved July 3, 2019, from https://www.drawdown.org/

6 EPA Climate Change Indicators: Greenhouse Gases. (2017, February 22). Retrieved from https://www.epa.gov/climate-indicators/greenhouse-gases

7 Financial Risk Management News Analysis. (2019, July 25). Retrieved from https://www.risk.net/; Glossary. (n.d.). Retrieved from http://www.risk.net/glossary

8 Frank, E. F., Witten, Hall, M. H., Eibe, H, I., & Mark. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). US: Morgan Kaufmann Publishers Inc.

9 Global warming of 1.5[degrees]C, an IPCC special report on the impacts of global warming of 1.5[degrees]C above pre-industrial levels, (website overview). (2019). *Library Journal, 144*(4), 40.

10 Guðbrandsdóttir, H. N., & Haraldsson, H. Ó. (2011). Predicting the price of EU ETS carbon credits. *Systems Engineering Procedia, 1*, 481-489. doi:10.1016/j.sepro.2011.08.070

11 Lackner, K. S., Brennan, S., Matter, J. M., Park, A., Wright, A., & Van Der Zwaan, B. (2012). The urgency of the development of CO_2 capture from ambient air. doi:10.7916/D8348W3G

12 McNall, S. G. (2011;2012;). *Rapid climate change: Causes, consequences, and solutions*. New York: Routledge. doi:10.4324/9780203834244

13 Satterthwaite, D. (2009). The implications of population growth and urbanization for climate change. *Environment and Urbanization, 21*(2), 545-567. doi:10.1177/0956247809344361

14 The Social Cost of Carbon. (2017, January 09). Retrieved from https://19january2017snapshot.epa.gov/climatechange/social-cost-carbon_.html

15 World Bank Open Data. (2015). Retrieved July 3, 2019, from https://data.worldbank.org/

16 Worldmeters Population by Year. (n.d.). Retrieved from https://www.worldometers.info/world-population/world-population-by-year/

17 (Ye, Xie, Zhang, & Hu, 2018)

Ye, F., Xie, X., Zhang, L., & Hu, X. (2018). An improved grey model and scenario analysis for carbon intensity forecasting in the pearl river delta region of china. *Energies, 11*(1), 91. doi:10.3390/en11010091

18 Zhu, B., & Wei, Y. (2013). *Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology* doi:6149/10.1016/j.omega.2012.06.005

19 UNdata Glossary. (n.d.). Retrieved from http://data.un.org/Glossary.aspx

20 What is Carbon Pricing? (n.d.). Retrieved from https://carbonpricingdashboard.worldbank.org/what-carbon-pricing

21 World Population by Year. (n.d.). Retrieved July 9, 2019, from https://www.worldometers.info/world-population/world-population-by-year/

22 Carbon Pricing Dashboard (n.d.). Retrieved from https://carbonpricingdashboard.worldbank.org/map_data

APPENDIX A

Table 13 Data Analysis Terms (UNdata Glossary), (Financial Risk Management News Analysis - Glossary, 2019)

Carbon Pricing Initiative Value [billion US$]	Cost of the initiative to be implemented and value in terms of CO_2 emission reduction
Certified Emission Reduction (CER)	The right to emit 650,000 tons of CO_2. CER is the technical term for the output of Clean Development Mechanism (CDM) projects, as defined by the Kyoto Protocol. A unit of greenhouse gas reductions that has been generated and certified under the provisions of Article 12 of the Kyoto Protocol, the CDM.
CO_2_Emissions_FF_Country ($GtCO_2$)	The world's countries emit vastly different amounts of heat-trapping gases into the atmosphere. Here's an estimate **carbon dioxide** emission from the combustion of coal, natural gas, oil and other fuels, including industrial waste and non-renewable municipal waste.
Cotation Assistée en Continu (CAC)	**CAC 40** is the French stock market index that tracks the **40** largest French stocks based on the Euronext Paris market capitalization. BREAKING DOWN **CAC 40 CAC 40** stands for Cotation Assistée en Continu, which translates to continuous assisted trading, and is used as a benchmark index for funds investing in the French stock market.
Country_Agriculture, hunting, forestry, fishing (ISIC A-B)	The Agriculture, Forestry, Fishing and Hunting sector comprises establishments primarily engaged in growing crops, raising animals, harvesting timber, and harvesting fish and other animals from a farm, ranch, or their natural habitats.
Country_Exports of goods and services	Exports of goods and services consist of sales, barter, or gifts or grants, of goods and services from residents to non-residents. The treatment of exports and imports in the SNA is generally identical with that in the balance of payments accounts as described in the Balance of Payments Manual.
Country_Final consumption expenditure	Final consumption expenditure consists of household final consumption expenditure, government final consumption expenditure and final consumption expenditure of NPISH's. Final consumption expenditure consists of expenditure incurred by resident institutional units on goods or services that are used for the direct satisfaction of individual needs or wants, or the collective needs of members of the community.
Country_Gross Domestic Product (GDP)	Gross domestic product is an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production (plus any taxes, and minus any subsidies, on products not included in the value of their outputs). The sum of the final uses of goods and services (all uses except intermediate consumption) measured in purchasers' prices, less the value of imports of goods and

	services, or the sum of primary incomes distributed by resident producer units.
Country_Manufacturing (ISIC D)	Manufacturing represents the economic activities of section D Manufacturing (see ISIC Rev 3.1).
Country_Mining, Manufacturing, Utilities (ISIC C-E)	Mining, manufacturing and utilities is an aggregation of economic activities of Section C Mining and quarrying, Section D Manufacturing and Section E Electricity, gas and water supply (see ISIC Rev 3.1).
Deutscher Aktienindex (DAX)	The DAX is a blue-chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange. Prices are taken from the Xetra trading venue. According to Deutsche Börse, the operator of Xetra, DAX measures the performance of the Prime Standard's 30 largest German companies in terms of order book volume and market capitalization. It is the equivalent of the FT 30 and the Dow Jones Industrial Average, and because of its small selection it does not necessarily represent the vitality of the economy as whole.
Gasoil	A middle distillate and form of heating oil used primarily in heating and air-conditioning systems. One of the most actively traded oil products, gasoil is the underlying in a key International Petroleum Exchange (IPE) futures contract. In refining terms, gasoil comes between fuel oil and the lighter products such as naphtha and gasoline. In its broader definition, it covers the oil products used for diesel automotive fuel and jet fuel.
GHG emissions covered [$MtCO_2e$]	The amount of CO_2 emissions addressed or reduced due to implemented initiative
Global Means_SurfaceTemp_Chg_12_Mth_ Avg	GISS measures the change in global surface temperatures relative to average temperatures from 1951 to 1980. Anomalies calculated for 2017 were **1.5 degrees F** (0.83 C) higher than the average temperatures for all the years in the 20th century.
Grey relational degrees	Information quantity and quality form a continuum from a total lack of information to complete information – from black through grey to white.
Grey system theory	A grey system means that a system in which part of information is known and part of information is unknown. It defines situations with no information as black, and those with perfect information as white.
Human development index (HDI)	A statistic composite index of life expectancy, education, and per capita income indicators, which are used to rank countries into four tiers of human development. A country scores a higher HDI when the lifespan is higher, the education level is higher, and the gross national income GNI (PPP) per capita is higher.

Information Gain	Information gain is defined as the entropy of the parent minus the weighted average of the entropy of the children that would result if you split that parent. At each node of a decision tree, the feature with the *largest information gain* is chosen for the split. The process is applied recursively from the root-node down and stops when a leaf node contains instances all having the same class or no gain (no need to split further).
Intergovernmental Panel on Climate Change (IPCC)	a group of scientists convened by the United Nations to guide world leaders on Climate Change
Portuguese Stock Index (PSI20)	The **PSI-20** is a benchmark stock market index of companies that trade on Euronext Lisbon, the main stock exchange of Portugal. The index tracks the prices of the twenty listings with the largest market capitalization and share turnover in the PSI Geral, the general stock market of the Lisbon exchange. It is one of the main national indices of the pan-European stock exchange group Euronext alongside Brussels' BEL20, Paris's CAC 40 and Amsterdam's AEX.
Price_rate_1_2019 (class)	Carbon tax or price charge for CO_2 emissions
Pruning (Decision Trees)	Technique in machine learning and search algorithms that reduces the size of decision trees by removing sections of the tree that provide little power to classify instances. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting. (Frank et al., 2011)
Savings from Initiative [billion US$]	Savings realize due to implemented initiative; seem in emission reduction
Supervised ML	The process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers; the algorithm iteratively makes predictions on the training data and is corrected by the teacher. **Supervised**: All data is labeled, and the algorithms learn to predict the output from the input data. (Brownlee, 2016)
Unsupervised ML	Unlike supervised learning, there is no correct answers and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data. **Unsupervised**: All data is unlabeled, and the algorithms learn to inherent structure from the input data. (Brownlee, 2016)
Year of abolishment	Year the initiative was ended
Year of implementation	Year the initiative was implemented