Supporting Environmental Decision MakingApplication of Machine Learning Techniques to Australia's Emissions

Acheampong, Alex O; Boateng, Emmanuel B.

Published in:
Applied Intelligent Decision Making in Machine Learning

DOI:
10.1201/9781003049548-9

Supporting Environmental Decision-making: Application of Machine learning techniques to Australia's emissions

Alex O. Acheampong
Newcastle Business School, University of Newcastle, Australia
Corresponding authors' email: alexopokuacheampong@gmail.com

Emmanuel B. Boateng
School of Health Sciences, University of Newcastle, Australia
Email: bannyboat16@gmail.com

Abstract

The development of a robust, high-quality and accurate model for forecasting carbon emission is a prerequisite for providing insights into environmental policies for achieving the Paris agreement on climate change. Research comparing the forecasting or predicting ability of different machine learning (ML) algorithms by focusing on carbon emissions and more specifically on Australian's carbon emissions remains scarce. This chapter, therefore, applied different ML techniques such as decision tree, random forest, extreme gradient boosting and support vector regression to model Australia's carbon emissions. The findings indicated that decision tree (DT) has a higher coefficient of determination (R^2) of 99.71%, followed by random forest (RF) with R^2 of 99.14%, extreme gradient boosting (XGBoost) with R^2 of 98.88% and support vector regression (SVR) with R^2 of 97.42%. In terms of accuracy, the tree-based models had the lowest errors. Overall, the DT model produced the most accurate predictions. On the other hand, the kernel-based model, Radial Basis Function (RBF)-SVR had comparatively higher errors. For computational efficiency, the DT and SVR models were more efficient than XGBoost and RF. Comparatively, the DT model ranked first among the other machine learning techniques utilised in this study based on the performance assessment metrics.

Keywords: Australia; Carbon emissions; Decision support; Machine learning

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

1. Introduction

This chapter sought to compare the forecasting/predicting ability of machine learning techniques such as decision tree, random forest, extreme gradient boosting and support vector regression by focusing on Australia's carbon emissions. Australia is one of the major carbon-emitting countries in the world. Ndevr Environmental (2019)[1] report indicates that Australia's carbon emissions for the year to March 2019 increased to approximately 561 million tonnes of CO_2 equivalent. The continued increase in carbon emissions, which is the primary greenhouse gas behind climate change, would cause a decline in agriculture yield, damage property and infrastructure facilities, increase commodity prices and financial instability (Climate, 2019). Australia has enacted numerous policies to mitigate carbon emissions; however, it is critical to have a futuristic understanding of Australia's carbon emissions.

Having a futuristic understanding of carbon emissions requires the use of advanced modelling or forecasting technique. In the existing literature, the majority of the studies have utilised classical statistical approach to model or forecast carbon emissions (see Acheampong & Boateng, 2019). However, given the chaotic, non-linearity and non-stationarity of variables for modelling carbon emissions, the classical statistical approach are not appropriate for modelling such a complex behaviour (Acheampong & Boateng, 2019; Gallo, Contò, & Fiore, 2014). Apart from the classical statistical approaches, other structural simulation approaches such as Computable General Equilibrium (CGE) models, Atmospheric Stabilisation Framework (ASF) model, Multiregional Approach for Resources and Industry Allocation (MARIA) model, and National Energy Modelling System (NEMS) have been employed by various organisation to forecast carbon emissions (see, Auffhammer, 2007; O'Neill & Desai, 2005; Zhao & Du, 2015). It is argued against these models that since the value parameters are fixed using personal judgements and calibrations, it may not be able to capture the behaviour of the real economy (cited in Zhao & Du, 2015). Further, Grey Model (GM) has been used a technique for forecasting carbon emissions, however, Yin and Tang, (2013) argue that the GM model and especially GM (1,1) works well with limited data. Various studies have comparatively analysed the prediction ability of these models with machine learning algorithms such as Artificial Neutral Network (ANN) have revealed that machine learning techniques forecast better than these models (Falat & Pancikova, 2015; Stamenković, Antanasijević, Ristić, Perić-Grujić, & Pocajt, 2015; Valipour, Banihabib, & Behbahani, 2013).

[1] https://ndevrenvironmental.com.au/tracking-2-degrees-fy2019-q3/

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

Recently, the advancement in machine learning techniques has played a critical role in decision-making. For instance, machine learning algorithms such as decision tree, random forest, extreme gradient boosting and support vector regression have played a critical role in modelling and forecasting. Although there have been studies evaluating the forecasting ability classical statistical models and machine learning techniques, research comparing the forecasting or predicting the ability of different machine learning algorithms by focusing on carbon emissions and more specifically on Australian's carbon emissions remains scarce. This warrant further empirical study. Therefore, this chapter aims to evaluate the forecasting/predicting ability of four machine learning techniques such as decision tree, random forest, extreme gradient boosting and support vector regression by focusing on Australia's carbon emissions.

This chapter contributes to the literature in the following ways: First, this study contributes to the literature by investigating the forecasting/predicting ability of four machine learning techniques such as decision tree, random forest, extreme gradient boosting and support vector regression by focusing on Australia's carbon emissions, which is rare in the literature. Second, this study employs macroeconomic variables such as population size, economic growth, energy consumption, trade, financial development, foreign direct investment and urbanisation, which are mostly used as inputs to model carbon emissions (see, Acheampong, 2019; Acheampong, Adams & Boateng, 2019; Adams & Acheampong, 2019; Acheampong, 2018). Third, to achieve robust and efficient estimates, this study uses high-frequency data. Finally, the outcome of this study will inform environmental policymakers the best machine learning technique appropriate for forecasting carbon emissions in Australia. The rest of the chapter is organised as follows. In Section 2, Methodology and data are presented in Section 2, while the results are discussed in Section 3. Conclusions and Policy implications are presented in Section 4.

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

2. *Data and methodology*

2.1. Data

To compare the performance accuracy of the decision tree, random forest, extreme gradient boosting and support vector regression, this study utilised a dataset for the period 1960-2018. Following Acheampong & Boateng (2019), we convert the data from annual dataset to a quarterly dataset using the quadratic sum approach. Therefore, quarterly data, which range between 1960Q1-2018Q4, was used for the analysis. The output and input variable used for our analysis is presented in Table 1. Economic growth, energy consumption, financial development, foreign direct investment, physical investment, population size, trade and urbanisation are used as the input variables for modelling carbon emissions[2]. All the variables were obtained from the World Bank (2019)[3].

Table 1. Variables used in this study

Variables	**Proxies**	**mean**	**sd**	**min**	**max**
Economic growth	GDP per capita (constant 2010 US$)	36690.75	11635.62	19245.41	56919.38
Energy use	Energy use (kg of oil equivalent per capita)	4805.486	808.1064	3063.554	5964.666
Population size	Population, total	1.70E+07	4115748	1.03E+07	2.50E+07
Financial development	Domestic credit to private sector (% of GDP)	64.06122	41.96032	17.65457	142.2841
Urbanisation	Urban population growth (annual %)	1.644195	0.569214	0.768615	3.571777
Trade openness	Trade (% of GDP)	34.12007	6.410204	24.79318	45.7979
Physical capital investment	Gross capital formation (% of GDP)	27.83713	2.839597	22.39053	33.68189
Foreign direct investment	Foreign direct investment, net inflows (% of GDP)	2.171749	1.544092	-3.61882	7.005444
Carbon emissions	CO2 emissions (kt)	252419.8	92195.4	88202.35	394792.9

2.2. Methodology

2.2.1. Decision tree

Decision trees (DTs) are non-parametric supervised learning techniques applied to regression and classification problems (Das, Naik, & Behera, 2020). These tree-based models are popular

[2] see Acheampong & Boateng (2019)
[3] https://data.worldbank.org/

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

and advantageous in handling smaller data sets than neural network models. Through a repetitive process of splitting, the regression trees can yield a set of rules which can be used for prediction (Tso & Yau, 2007; Das, Naik, & Behera, 2020). The splitting process divides the sample into two or more homogeneous sets derived from the most important differentiator among the input variables. In the case of classification problems, metrics such as cross-entropy or Gini index are used to decide strategic splits for decision trees (Xu *et al.* 2005; Das et al., 2020). For regression problems, DTs normally use the mean squared error (MSE) criterion in splitting a node into two or more sub-nodes. That is, on each subset of data, the algorithm computes an MSE value and the tree with least MSE is selected as a point of split. The concluding outcome comprises decision nodes and leaf nodes (Das, Naik, & Behera, 2020a; 2020b). By contrast to black-box models such as deep learning models, DTs are easy to understand and interpret because its rules can be visualised. These algorithms have been applied and attained successes in many fields due to their efficiency and interpretability (Tsai & Chiou, 2009; Wu, 2009).

2.2.2. Random forest

Random forest (RF) is an ML algorithm that combines several DT models to effectively classify or predict an outcome (Breiman, 2001; Das, Naik, & Behera, 2020a; 2020c). This combination process also termed as bootstrap aggregation or bagging involves training each DT with a distinct set of observations through sample replacement. Samples which do not end up in a subset of training data during bagging is included with other subsets called "out-of-bag" (Rodriguez-Galiano et al., 2015). Bagging minimises the variance of the base learner, however, has minimal influence on the bias (Rodriguez-Galiano et al., 2015). Basically, on each node, there is a random selection of variables out of all possible variables, then the best split among the selected variables is determined based on the lowest MSE. The final prediction is derived using an ensembling technique, which is, averaging the predictions of the previous regression trees (Das, Naik, & Behera, 2020). Due to the averaging of several trees, there is a considerably lower risk of overfitting (Breiman, 2001). The random sampling of training observations and random subsets of candidate variables for splitting nodes is a clear distinguishing factor of RF from DT.

2.2.3. Extreme gradient boosting

Extreme gradient boosting (XGBoost) is a scalable end-to-end tree boosting algorithm (Chen & Guestrin, 2016). It can be applied to regression, ranking and classification problems. This algorithm performs parallel tree learning using a novel sparsity-aware system (Chen &

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

Guestrin, 2016). XGBoost employs the second-order Taylor expansion to approximate the loss function. This model has been known to outperform other ML models, and its successes can be witnessed in numerous machine learning and data-driven competitions such as Kaggle (Dey et al. 2019; Rout et al., 2020). The fundamental factor behind the triumphs of XGBoost is its scalability in all circumstances (Chen & Guestrin, 2016). That is, using an optimal amount of resources, the algorithm yields state-of-the-art results when solving a wide range of problems. The implementation of this model is, thus, influenced by its high speed and performance.

2.2.4. Support vector regression

Support vector regression (SVR) is a type of support vector machine used for regression purposes and hence handles continuous values (Das, Naik, & Behera, 2020, Dey et al., 2019). It follows the same principles as the support vector classification, however with minimal modifications. The most important distinguishing factor between a simple linear regression model and SVR is that a linear regression model attempts to reduce the error rate while SVR tries to fit the error within a designated threshold. This kernel-based model has advantages in high dimensional spaces since its optimisation does not rely on the dimensionality of the input area (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997; Das et al., 2018). Moreover, SVR is a non-parametric tool and does not rely on distributions of the primary input and output variables. Although less popular than support vector classification (SVC), SVR has proven to be an effective technique in solving real-world scale problems (Awad & Khanna, 2015).

2.3. Data division and experiment environment

We compared the accuracies and computational costs of these four ML algorithms. Data used for training and validating the models were standardised to eliminate instances of one variable dominating the other (Boateng, Pillay, & Davis, 2019) since the variables used in this study have different units (Acheampong & Boateng, 2019; Bannor & Acheampong, 2019). 80% (188 quarters) of the data were used to train each model while the remaining 20% (48 quarters) were used to validate the models. Similar data proportions were used by Morano *et al.* (2015); Lam *et al.* (2008) Acheampong & Boateng (2019); and Bannor & Acheampong (2019). For hardware and software environment, we used an Intel i5-2520M (4) at 3.2GHz CPU, 8GB memory operating on Ubuntu 18.04.2 LTS. Two GPUs were used, Intel 2nd Generation Core Proce, and NVIDIA GeForce GTX 1050Ti with 8GB memory. We used Spyder (Python 3.6.7) to write and execute the programming codes.

2.4. Hyperparameters optimisation

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

As the goal of this study is to evaluate predictive ML models, there is the need to develop optimal models apt for comparison purposes. We performed a grid-search with 10-fold cross-validation on the hyperparameters of each model. This technique shuffles and resamples the training data into 10 equal folds, fits the model on a combination of one set of hyperparameters on nine folds and tests the model on the remaining fold (Bannor & Acheampong, 2019). A carefully tuned model is at lower risks of underfitting and overfitting problems. The best score function returns the combination of hyperparameters and associated arguments suitable to develop the model. For all models, we assessed their mean cross-validation scores (MCVs). The highest MCV is used as the basis to select the ideal combination of hyperparameters.

2.4.1. Parameter tuning for the DT, RF, and XGBoost algorithms

Upon prior experimentations, certain hyperparameters were deemed to influence the performance of the models significantly and hence used in tuning the ML algorithms. In the case of the DT algorithm, five different maximum leaf nodes (None, 2, 3, 5, and 7), five maximum tree depths (1, 3, 5, 7, and 9), five minimum samples for a leaf node (1, 3, 5, 7, and 9) over the 10-folds of data results in 1250 models. We use a random state of zero for the DT regressor. For the RF algorithm, five numbers of estimators/trees (100, 150, 200, 250, and 300), five maximum tree depths (1, 3, 5, 7, and 9), and five minimum samples for a leaf node (1, 3, 5, 7, and 9) were specified. This also led to 1250 models. Hyperparameters arguments of the XGBoost such as number of estimators (100, 150, 200, 250, and 300), number of maximum tree depths (1, 3, 5, 7, and 9) and learning rates (0.01, 0.05, 0.1, 0.2, and 0.3) were also tuned totalling 1250 models.

2.4.2. Parameter tuning for the SVR algorithm

For the SVR algorithm, six penalty "*C*" parameter arguments (10.0, 50.0, 100.0, 1000.0, 1050.0 and 1100.0), seven different gamma values (0.0005, 0.0001, 0.001, 0.01, 0.1, 1.0, and 10.0), and three types of kernels (radial basis function, linear function, and polynomial function) were experimented over 10-folds of the training data set using the grid-search framework. In all, 1260 models were developed for the SVR. After the grid-search with 10-fold cross-validation exercise, we selected the ideal hyperparameter arguments for each algorithm based on their MCV scores.

2.5. Performance metrics

We assessed the accuracies of each algorithm in predicting the 48 carbon emission data points in the test data set. By evaluating the deviations between the predicted and actual emissions,

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

models with lower errors were ranked high in terms of accuracy. The root-mean-square error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) were used in comparing the levels of deviations among the four ML algorithms. We also assessed the computational efficiency of the four models. Particularly, the elapsed time taken during the grid-search with 10-fold cross-validation process on each algorithm was measured.

3. Results and discussion

3.1. Development and validation of the DT model

After performing the grid-search with 10-fold cross-validation, the DT model with the parameter combination of a maximum tree depth of 9, maximum leaf nodes as "None", and minimum samples for a leaf node as 1 had the highest MCV score of 0.975899. These parameters and their arguments were used in configuring the DT model, then train and validate it. This relationship between the hyperparameters and the MCV scores during the cross-validation process is depicted in Figure 1. It can be observed that the higher the depth of the tree, the higher the maximum leaf nodes grows and hence a more accurate DT model. That is, the "None" argument of the maximum leaf nodes parameter suggest an unlimited number of nodes relative to the decrease in errors. On the other hand, a higher tree depth corresponds well with few minimum samples required for a leaf node.

Sections of the validated DT tree model are shown in Figures 2a and 2b. Negligible MSEs can be observed at each internal and leaf node, and hence suitable to compete with the other three ML models. The variables X_0 to X_7 within the internal nodes tallies with the features in the input data set; thus, the first feature "credit to private sector" is X_0, "GDP" is X_1, "physical investment" is X_2, "population" is X_3, "trade" is X_4, "urbanisation" is X_5, "energy use" is X_6, and "FDI" is X_7.

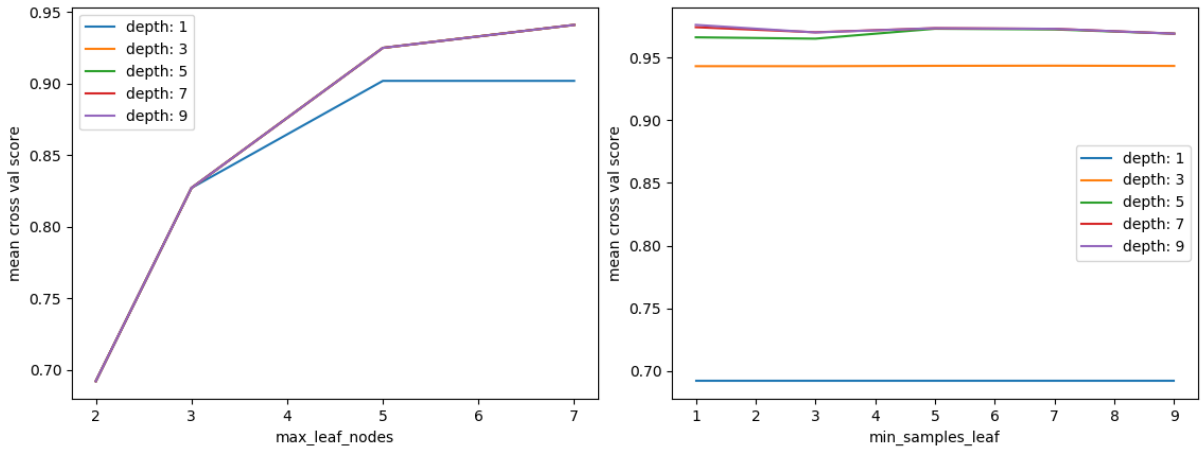

Fig. 1. Hyperparameters optimisation for DT models

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

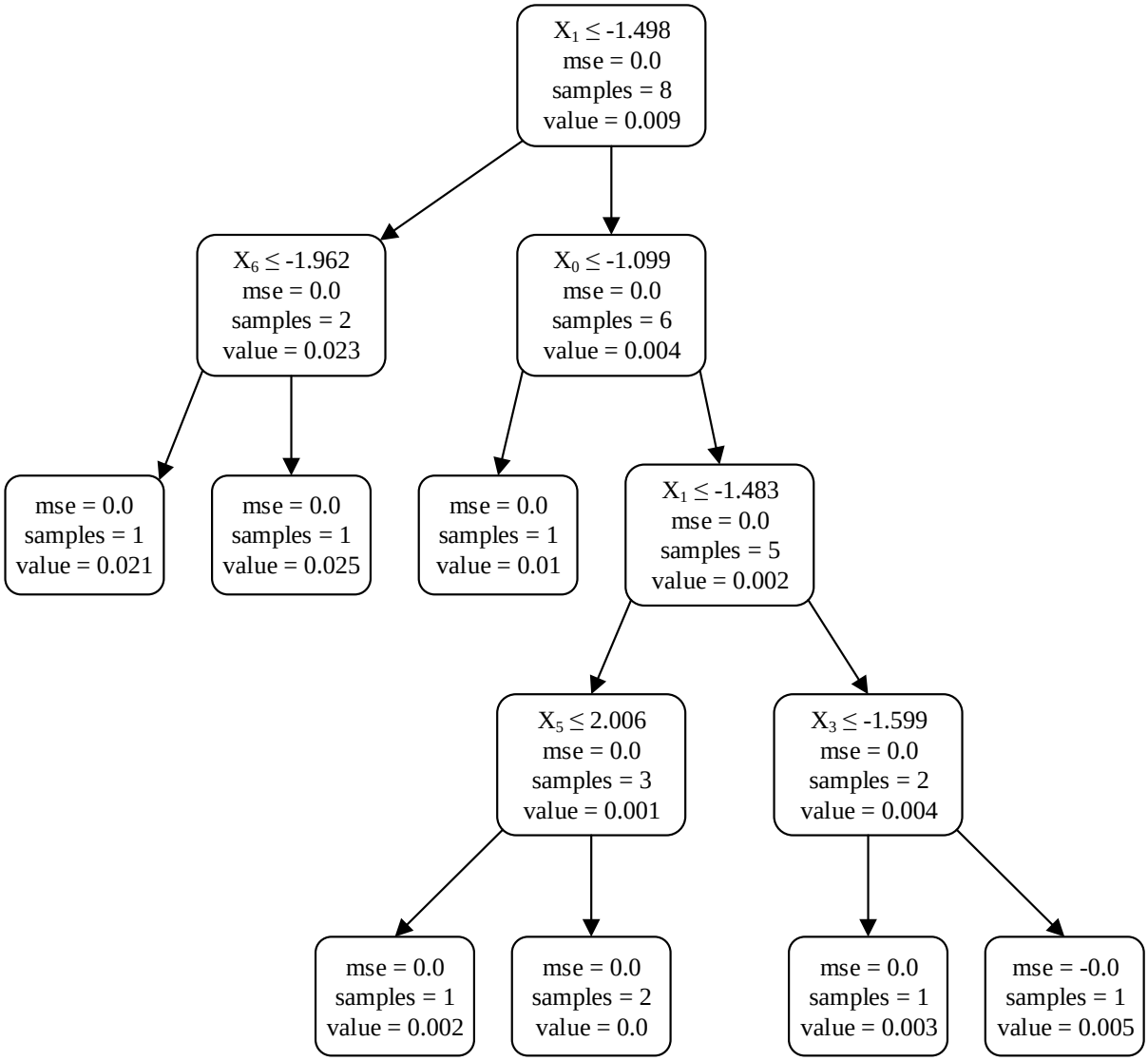

Fig. 2a. Section of DT model plot

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

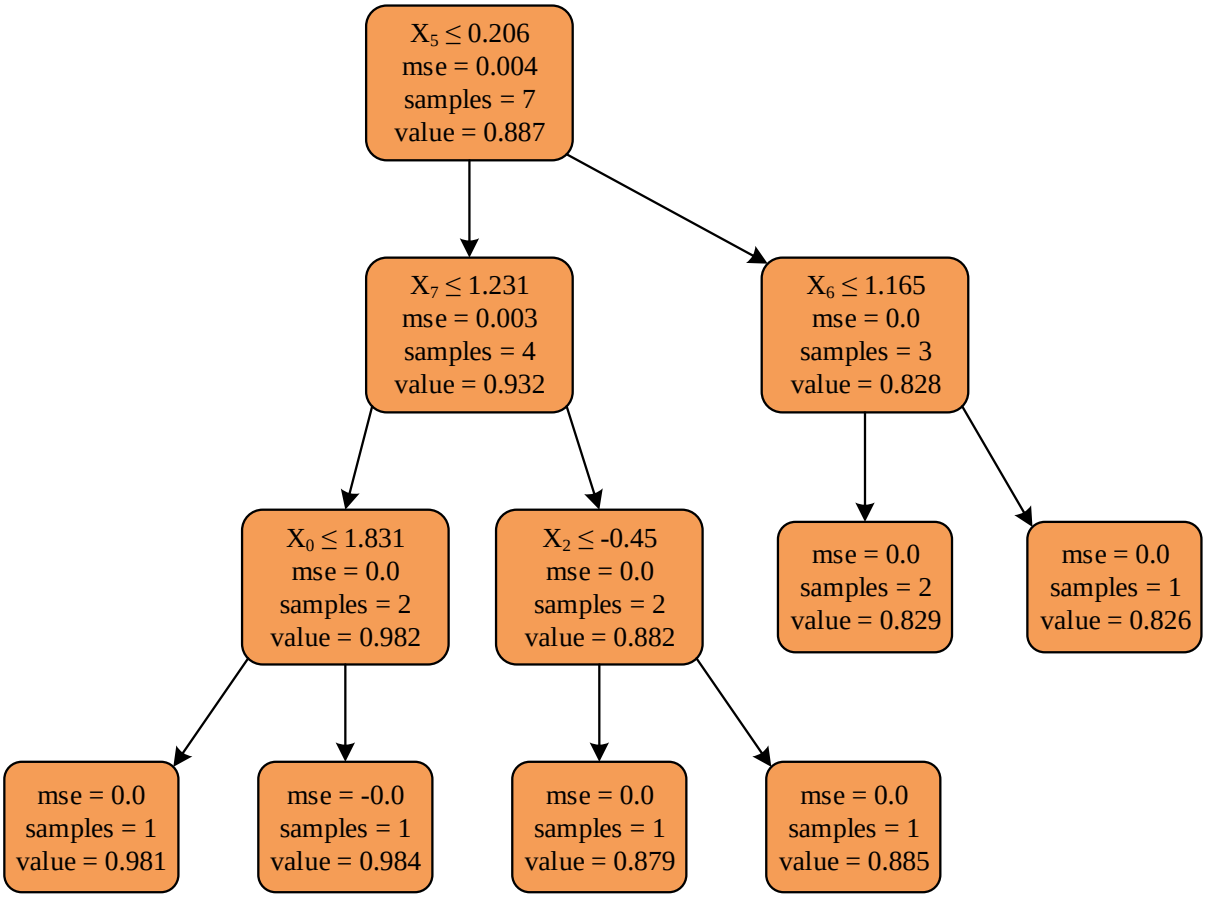

Fig. 2b. Section of DT model plot

3.2. Development and validation of the RF model

The RF model with the parameter combination of a maximum tree depth of 9, minimum samples for a leaf node as 1, and a number of estimators/trees as 250 had the highest MCV score of 0.984858 attained after evaluating all 1250 RF models from the grid-search with 10-fold cross-validation process. Corroborating the findings from the DT model, higher tree depths is inversely related to minimum samples required for a leaf node. For instance, in Figure 3, the maximum number of tree depth of 9 corresponds to a minimum sample of 1 required for a leaf node.

Figure 4 shows the DT model configured with the best hyperparameter arguments per the grid-search, then trained with 188 quarters of emissions data and validated on 48 quarters. Similar to the DT model, the RF model's (Figure 4) variables of X_0 to X_7 in the root and internal nodes corresponds to the features "credit to the private sector" to "FDI" respectively. In Figure 4, these features are considered in constructing questions while accessing a random set of training observations. The number of samples on each internal node are also depicted, and associated errors on the predictions or decisions on each node are evaluated with the MSE

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

criterion. Our RF model consists of 250 DTs combined into a single predictive model. We plot only the 250th DT at a maximum depth of 2 due to brevity. Using the wisdom of the crowd, the resulting outcomes or predictions of the RF model are made in consultation with other DTs by averaging out their predictions.

In Figure 4, the first node is the root node, while the remaining nodes are the internal nodes. As this tree is only shown at a depth of 2, the leaf nodes are hard to visualise (due to brevity) as the complete maximum depth of a tree is 9 in this study. For the 250th tree at a maximum depth of 2, only three features were important according to the model and hence used to making decisions. These three were GDP (X_1), population (X_3), and energy use (X_6).

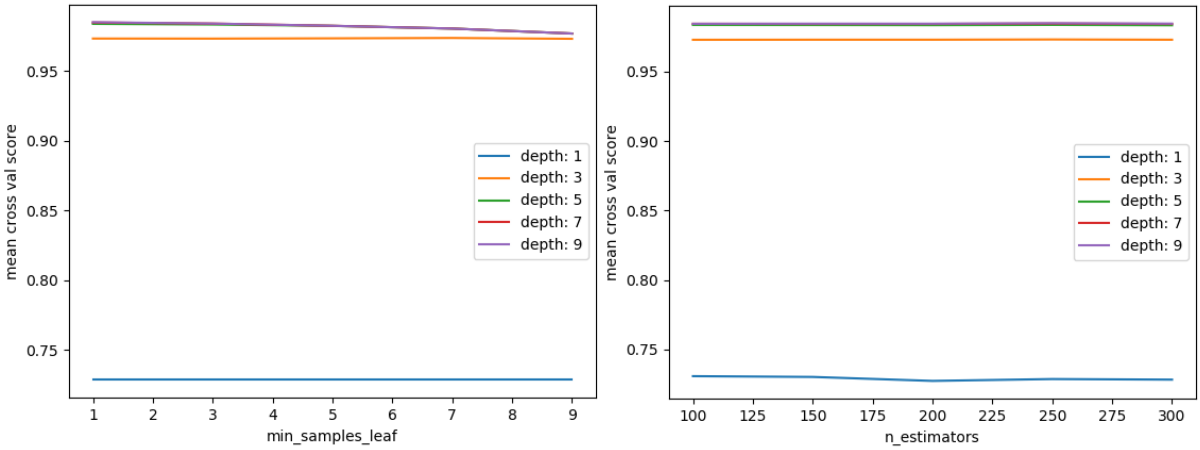

Fig. 3. Hyperparameters optimisation for RF models

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

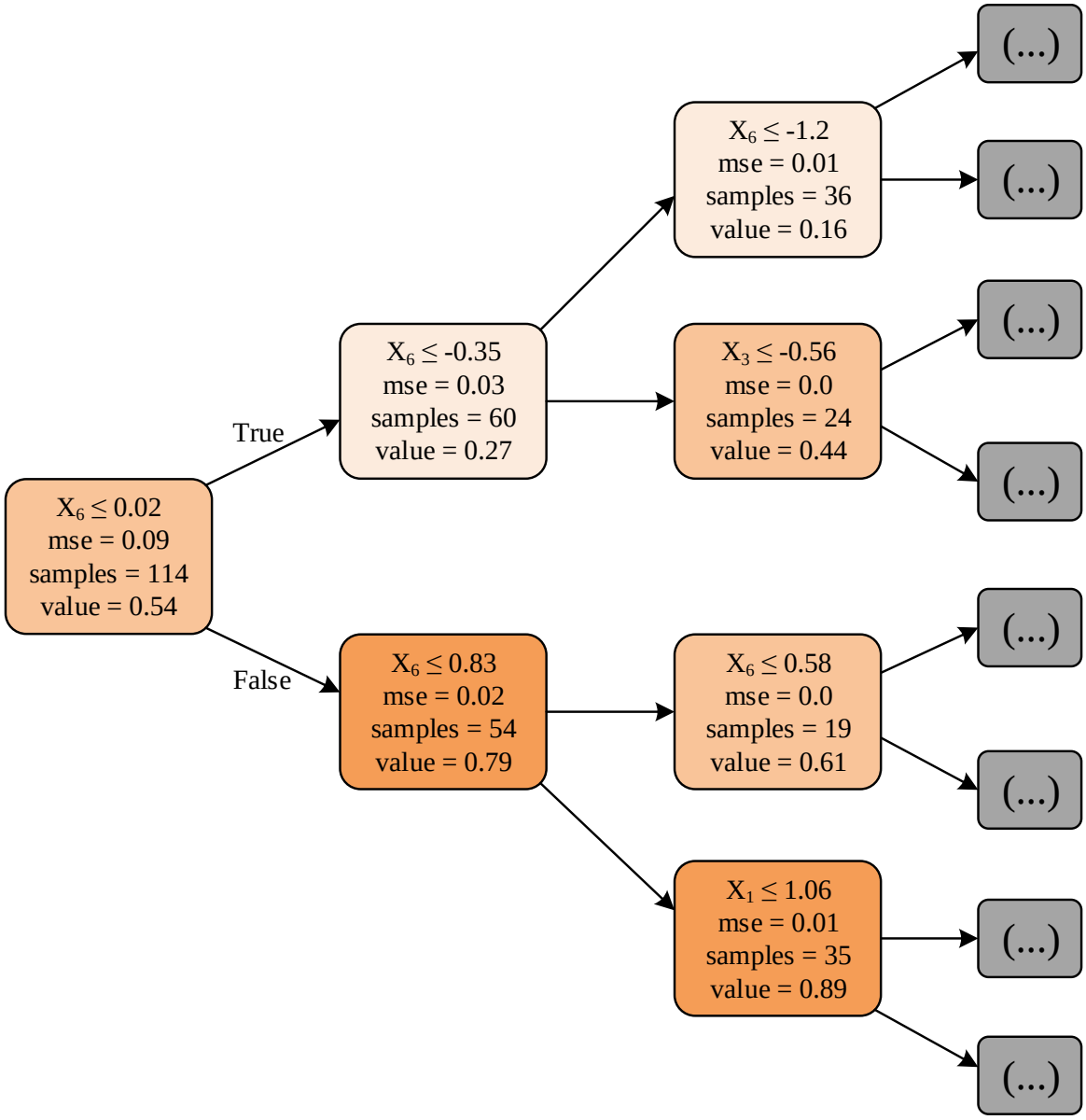

Fig. 4. RF model plot of 250th tree at a maximum depth of 2

3.3. Development and validation of the XGBoost model

Next, MCV scores of 1250 XGBoost models were evaluated after the grid-search with 10-fold cross-validation. The model with the hyperparameter combination of a 0.05 learning rate, maximum tree depth of 3, and a number of estimators/trees as 300 attained the highest MCV score of 0.985607. Figure 5 shows that irrespective of the maximum number of tree depth, higher MCV scores were associated with the 0.05 learning rate. This scenario is evident as an upwards kink can be observed at the 0.05 learning rate. These ideal parameters as determined by the grid-search framework were used to build, calibrate, and test the model (Figure 6).

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

As shown in Figure 6, we plot the 300th tree of the XGBoost model. While "X_n" is used as a feature in the DF and RF models, "fn" is used for XGBoost. Specifically, X_0 to X_7 in the DT and RF models correspond to f0 to f7 in the XGBoost model. At the 300th tree, four features (GDP, physical investment, trade, and energy use) were deemed important by the model and hence used in making the predictions. The important features used to make the decisions often vary from tree to tree and in other instances, some trees may have the same features. It can also be observed that the influence of the maximum tree depth of 3 matches the 3 layers of internal nodes in Figure 6. The questions or conditions on the internal nodes suggest which data could sieve through. The scores on the leaf nodes are also termed as weights.

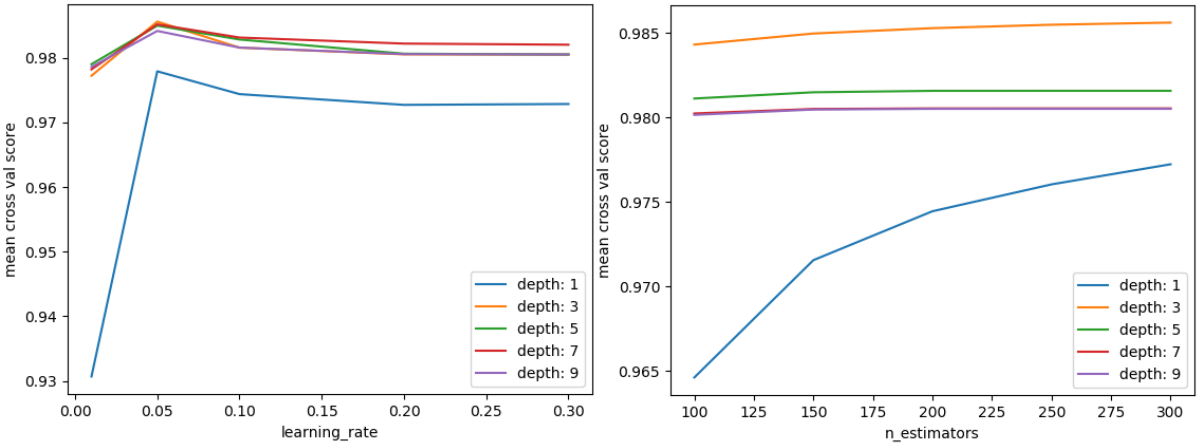

Fig. 5. Hyperparameters optimisation for XGBoost models

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

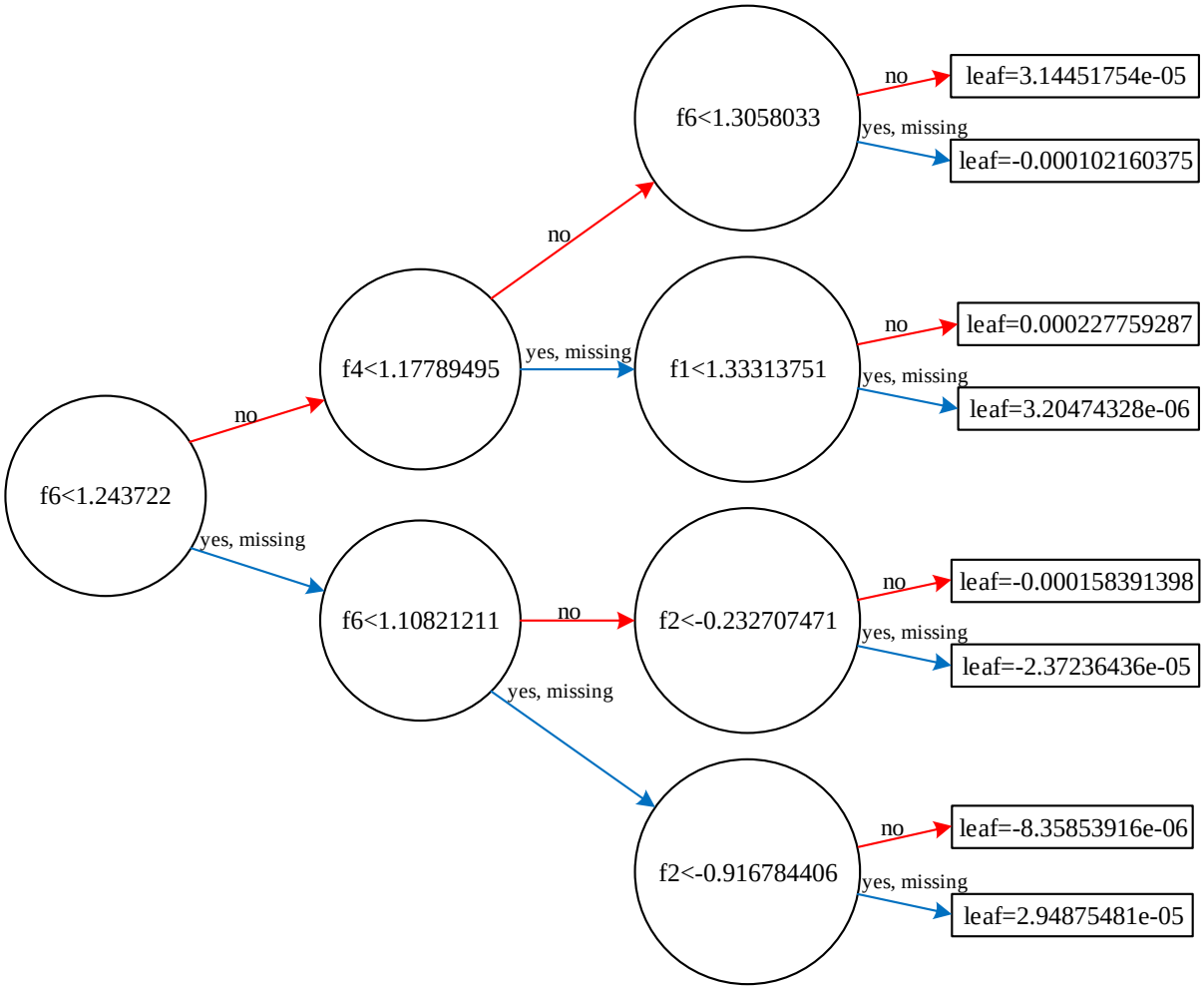

Fig. 6. XGBoost model plot at 300th tree

3.4. Development and validation of the SVR model

1260 SVR models were developed after performing the grid-search with 10-fold cross-validation. The model with hyperparameters of a *C* of 50.0, gamma of 0.001, and an RBF kernel reached the highest MCV score of 0.967804. Considering the difficulties of illustrating more than 3 features in a 3D chart, we show separately for each kernel the MCV scores attained in association with other hyperparameters (Figure 7). During the grid-search process, it can be observed that an increase in the gamma does not have a significant influence on the linear and polynomial SVR models.

On the other hand, a significant inverse relationship is seen between the gamma and the MCV score of the RBF models. This is because the gamma regulates the influence of new features on the decision boundary. Hence, a too large gamma could result in overfitting and no amount of *C* could avoid it. In order for the models to effectively predict accurately, the gamma is lowered to an optimal level to capture the complexity of the data. Figure 8 shows the trained and validated RBF-SVR model which was configured with the ideal parameters derived from

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

the grid-search process. The SVR model shows promising results as the best line of fit accurately passes through/near the test data points.

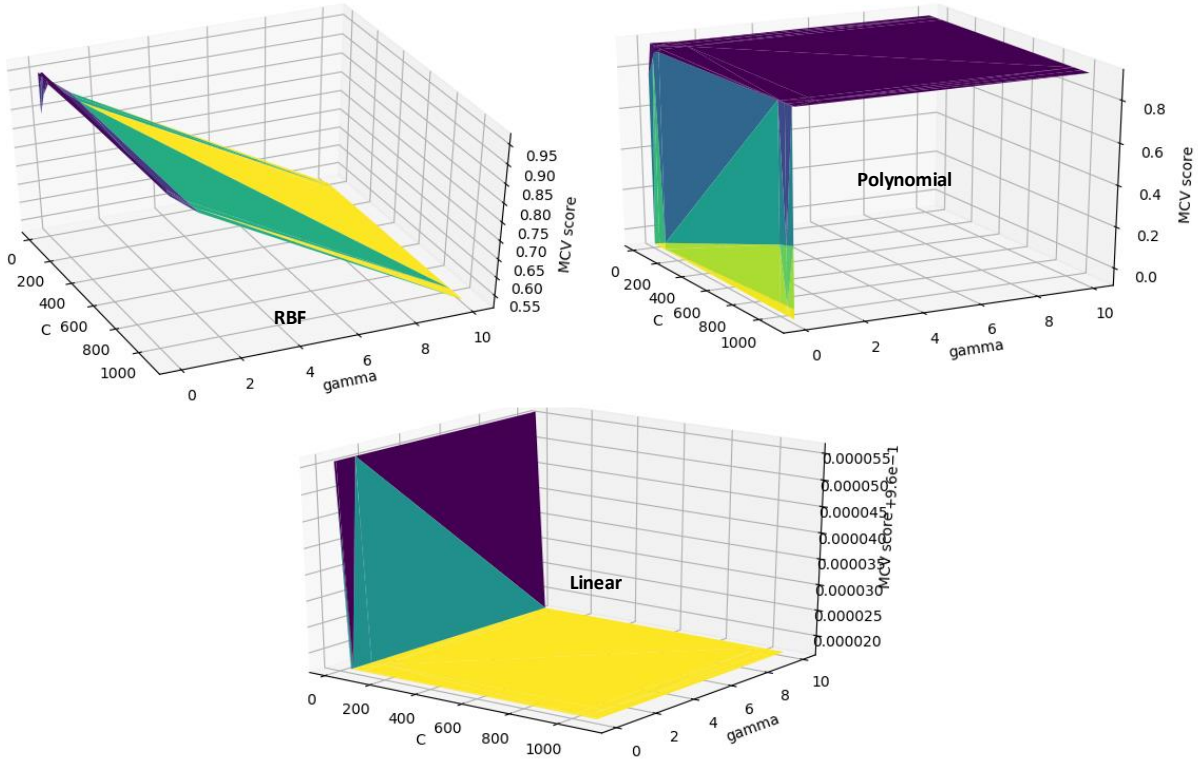

Fig. 7. 3D grid-search results for SVR models

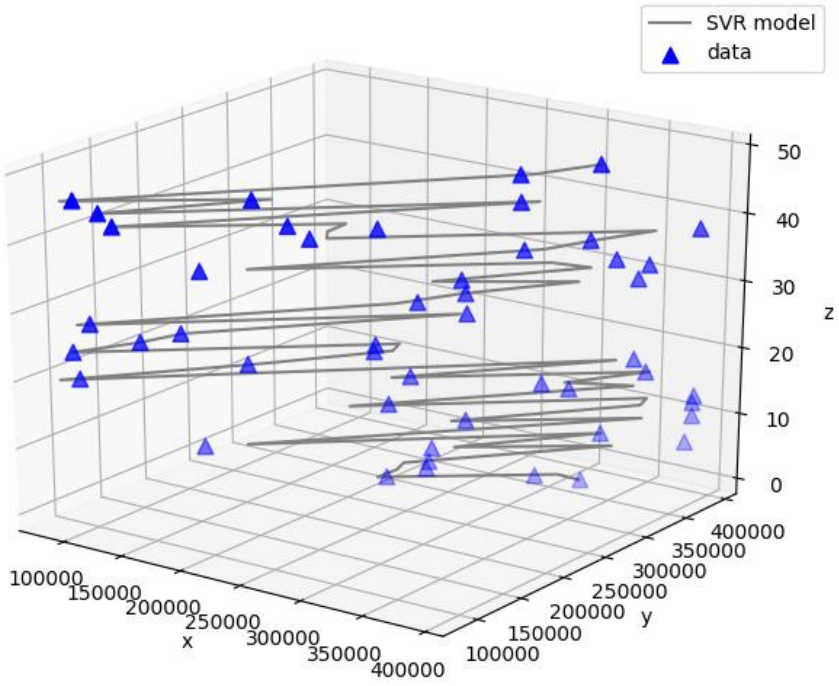

Fig. 8. SVR model in 3D view

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

3.5. Performance evaluation of model performance

We compared the amount of observed variance that each validated model explained (Figure 9). The DT model explained almost 100% of the observed variance, putting its' R^2 at 99.71%. And hence having its data points almost perfectly fitted on the regression line. The RF model followed closely with 99.14%. The XGBoost model also accounted for 98.88% of the observed variance around its mean. Though the SVR model had the lowest R^2 in this study, its ability to explain 97.42% of the observed variance only leaves less than 3% of the observed variance accounted for by the model.

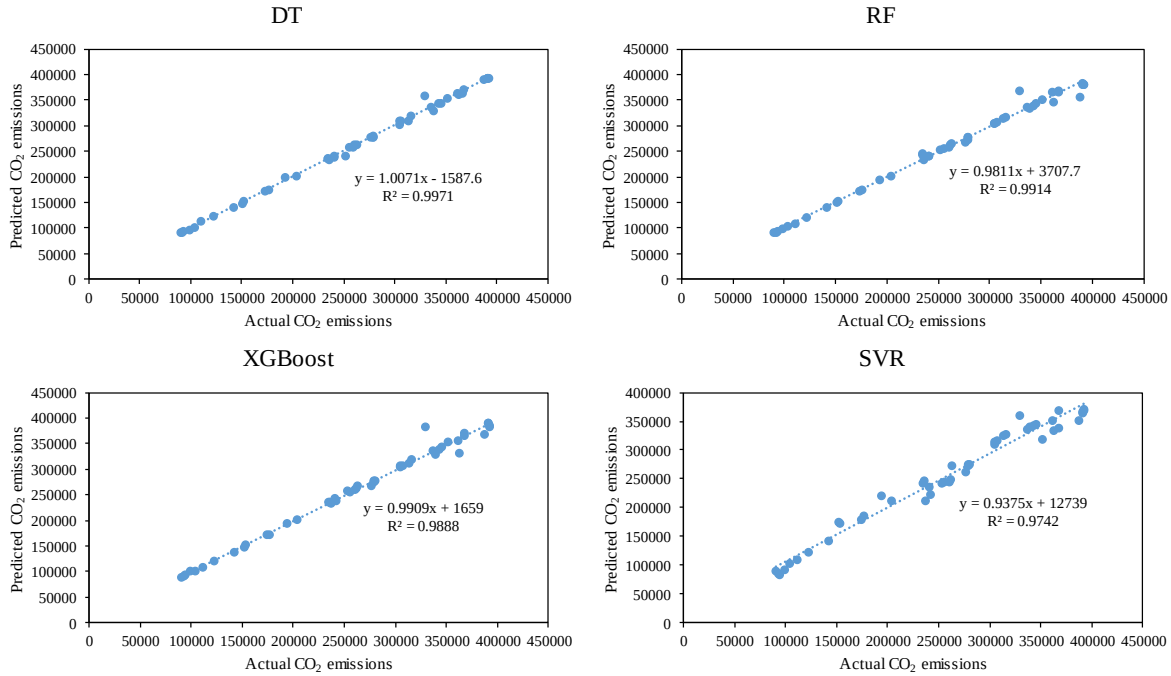

Fig. 9. Scatter diagram of actual and predicted CO_2 emissions

In terms of accuracy, the tree-based models had the lowest MAPE and RMSE, as shown in Table 2. In all, the DT model produced the most accurate predictions. On the other hand, the kernel-based model, RBF-SVR had comparatively higher errors. This could be observed in Figure 8 as the SVR model hardly passed through/near few of the test data points. For computational efficiency. The DT and SVR models were more efficient than the other two models. Overall, the DT model ranked top in all spheres of performance assessment in this study.

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

Table 2. Performance evaluation of ML models

Model	MAPE*	RMSE†	R^2	Elapsed time	Ranking	
					Accuracy	Time
DT	1.1076	5145.3328	0.9971	2.2s for 1250 fits	1st	1st
RF	1.4812	8834.8885	0.9914	1.9min for 1250 fits	2nd	4th
XGBoost	1.6228	9948.4937	0.9888	39.2s for 1250 fits	3rd	3rd
SVR	4.8045	15823.5338	0.9742	2.6s for 1260 fits	4th	2nd

†CO_2 emissions (metric tons per capita); *%

4. Concluding remarks

The development of a robust, high-quality and accurate model for forecasting carbon emission is a prerequisite for providing insights into environmental policies for achieving the Paris agreement on climate change. Research comparing the forecasting or predicting the ability of different machine learning algorithms by focusing on carbon emissions and more specifically on Australian's carbon emissions remains scarce. This warrants further research. This chapter, therefore, applied different machine learning (ML) techniques such as decision tree, random forest, extreme gradient boosting and support vector regression to model Australia's carbon emissions. The findings indicated that decision tree (DT) has a higher coefficient of determination (R^2) of 99.71%, followed by random forest (RF) with R^2 of 99.14%, extreme gradient boosting (XGBoost) with R^2 of 98.88% and support vector regression (SVR) with R^2 of 97.42%. In terms of accuracy, the tree-based models had the lowest MAPE and RMSE. Overall, the DT model produced the most accurate predictions. On the other hand, the kernel-based model, RBF-SVR had comparatively higher errors. For computational efficiency, the DT and SVR models were more efficient than XGBoost and RF. Comparatively, the DT model ranked first among the other ML techniques utilised in this study based on the performance assessment metrics. The implication is that for accurate prediction and effective planning of policies to mitigate climate change, Australia's environmental planners should incorporate a decision tree as one of the techniques for modelling and forecasting of carbon emissions. Some ML algorithms could have performed better in terms of their computational efficiency; however, the combination and choice of certain parameters for such algorithms during the grid-search process can be computationally laborious in general and hence at a marginal extent, could limit our results in those aspects. Other robust and novel ML algorithms could be employed to explain 100% variance in the dataset. As this study was for regression, future research could investigate for classification problems pertaining to Australia's carbon emissions.

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

References

Acheampong, A. O. (2018). Economic growth, CO2 emissions and energy consumption: What causes what and where? Energy Economics, 74, 677-692. doi:https://doi.org/10.1016/j.eneco.2018.07.022

Acheampong, A. O. (2019). Modelling for insight: Does financial development improve environmental quality? Energy Economics, 83, 156-179. doi:https://doi.org/10.1016/j.eneco.2019.06.025

Acheampong, A. O., & Boateng, E. B. (2019). Modelling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production, 225*, 833-856. doi:https://doi.org/10.1016/j.jclepro.2019.03.352

Acheampong, A. O., Adams, S., & Boateng, E. (2019). Do globalization and renewable energy contribute to carbon emissions mitigation in Sub-Saharan Africa? Science of The Total Environment, 677, 436-446. doi:https://doi.org/10.1016/j.scitotenv.2019.04.353

Adams, S., & Acheampong, A. O. (2019). Reducing carbon emissions: The role of renewable energy and democracy. Journal of Cleaner Production, 240, 118245. doi:https://doi.org/10.1016/j.jclepro.2019.118245

Auffhammer, M. (2007). The rationality of EIA forecasts under symmetric and asymmetric loss. Resource and Energy Economics, 29(2), 102-121.

Awad, M., & Khanna, R. (2015). Support Vector Regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (pp. 67-80). Berkeley, CA: Apress.

Bannor, B. E., & Acheampong, A., O. (2019). Deploying artificial neural networks for modeling energy demand: international evidence. *International Journal of Energy Sector Management, ahead-of-print*(ahead-of-print). doi:10.1108/IJESM-06-2019-0008

Boateng, E. B., Pillay, M., & Davis, P. (2019). Predicting the Level of Safety Performance Using an Artificial Neural Network. In T. Ahram, W. Karwowski, & R. Taiar (Eds.), *Human Systems Engineering and Design* (pp. 705-710). Cham: Springer International Publishing.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*.

Climate Council. (2019). *Compound cost: How climate change is damaging Australia's economy*. Retrieved from

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

Das, H., Naik, B., & Behera, H. S. (2018). Classification of diabetes mellitus disease (DMD): a data mining (DM) approach. In Progress in computing, analytics and networking (pp. 539-549). Springer, Singapore.

Das, H., Naik, B., & Behera, H. S. (2020a). An Experimental Analysis of Machine Learning Classification Algorithms on Biomedical Data. In Proceedings of the 2nd International Conference on Communication, Devices and Computing (pp. 525-539). Springer, Singapore.

Das, H., Naik, B., & Behera, H. S. (2020b). A Hybrid Neuro-Fuzzy and Feature Reduction Model for Classification. Advances in Fuzzy Systems, 2020.

Das, H., Naik, B., & Behera, H. S. (2020c). Medical disease analysis using neuro-fuzzy with feature extraction model for classification. Informatics in Medicine Unlocked, 18, 100288.

Das, H., Naik, B., Behera, H. S., Jaiswal, S., Mahato, P., & Rout, M. (2020). Biomedical data analysis using neuro-fuzzy model with post-feature reduction. Journal of King Saud University-Computer and Information Sciences.

Dey, N., Das, H., Naik, B., & Behera, H. S. (Eds.). (2019). Big data analytics for intelligent healthcare management. Academic Press.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). *Support vector regression machines.* Paper presented at the Advances in neural information processing systems.

Falat, L., & Pancikova, L. (2015). Quantitative Modelling in Economics with Advanced Artificial Neural Networks. Procedia Economics and Finance, 34, 194-201. doi:https://doi.org/10.1016/S2212-5671(15)01619-6

Gallo, C., Contò, F., & Fiore, M. (2014). A neural network model for forecasting CO2 Emission. *Agris on-line Papers in Economics and Informatics*, *6*(2), 31.

Lam, K. C., Yu, C. Y., & Lam, K. Y. (2008). An artificial neural network and entropy model for residential property price forecasting in Hong Kong. Journal of Property Research, 25(4), 321-342.

Morano, P., Tajani, F., & Torre, C. M. (2015). Artificial intelligence in property valuations an application of artificial neural networks to housing appraisal. Advances in Environmental Science and Energy, Planning, 23-29.

Ndevr Environmental (2019). Tracking 2019 Q3 Report. Retrived from Ndevr Environmental (2019). Tracking 2019 Q3 Report. Retrived from https://ndevrenvironmental.com.au/tracking-2-degrees-fy2019-q3/

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9

O'Neill, B. C., & Desai, M. (2005). Accuracy of past projections of US energy consumption. Energy Policy, 33(8), 979-993.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804-818.

Rout, M., Jena, A. K., Rout, J. K., & Das, H. (2020). Teaching–Learning Optimization Based Cascaded Low-Complexity Neural Network Model for Exchange Rates Forecasting. In Smart Intelligent Computing and Applications (pp. 635-645). Springer, Singapore.

Stamenković, L. J., Antanasijević, D. Z., Ristić, M. Đ., Perić-Grujić, A. A., & Pocajt, V. V. (2015). Modeling of methane emissions using artificial neural network approach. Journal of the Serbian Chemical Society, 80(3), 421-433.

Tsai, C.-F., & Chiou, Y.-J. (2009). Earnings management prediction: A pilot study of combining neural networks and decision trees. *Expert Systems with Applications, 36*(3), 7183-7191.

Tso, G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, 32(9), 1761-1768.

Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476, 433-441. doi:https://doi.org/10.1016/j.jhydrol.2012.11.017

World Bank. (2019). World development indicators. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators#

Wu, D. (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications, 36*(5), 9105-9112.

Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. Remote Sensing of Environment, 97(3), 322-336.

Yin, M.-S., & Tang, H.-W. V. (2013). On the fit and forecasting performance of grey prediction models for China's labor formation. Mathematical and Computer Modelling, 57(3), 357-365. doi:https://doi.org/10.1016/j.mcm.2012.06.013

Zhao, X., & Du, D. (2015). Forecasting carbon dioxide emissions. *Journal of Environmental Management, 160*, 39-44. doi:https://doi.org/10.1016/j.jenvman.2015.06.002

This is an Accepted Manuscript of a book chapter published by Routledge in H. Das, J. K. Rout, S. C. Moharana, & N. Dey (Eds.), Applied Intelligent Decision Making in Machine Learning on 19 Nov 2020, available online: https://doi.org/10.1201/9781003049548-9