International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-08, Issue-01, APR 2022

Predicting CO2 Emission Using Machine Learning

[1]Prof. Swapnil Wani,[2]Mr. Akash Akhilesh Yadav, [3]Mr.Mihir Mukesh Panchal, [4]Mr. Prashant Vinod Pandey

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. ShivajiraosS.Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India. [1]*swapnilwani27@gmail.com*, [2]*akki15yad@gmail.com*, [3]*mihirpanchaliaf7@gmail.com*,[4]*pandeyprashant953@gmail.com*

Abstract- The Random Forest & SVM are put forward to estimate the outlay of CO2 outpouring. Energy expenditure, such as power and coal energy, is the cause of the aggressive growth in CO2 emissions. The aim is to track co2 emissions gleaned from electrical energy and coal utilized in the manufacturing process. In setup to train and test the model, the statistics on electrical and energy were obtained. The statistics were separated into 60% training data and 40% testing data. To find the distinguished parameters of both model's, trial and error approach was used. The minimal error of the model reflects a more precise estimation while using RMSE to measure the model's error.

Keywords – CO2 Emmission, Machine Learning.

I. INTRODUCTION

Over the years, CO2 emissions have increased substantially in a few emergingnations while falling in big industrialized ones. Human activities like as the ignition of fuel, coal, and gas, as well as deforestation, are the major sources of rising co2 levels in the atmosphere. Continuous efforts are undertaken to get to use fewer fossil fuels which contribute more than 70% of total CO2 Emissions made across theglobe.

Countries that are already developed opted for greener energy sources and have invested tremendously in making it even greener. But along the same line developed countries are emitting even higher as compared to previous years majorly due to the increased industrial revolution lately.

Efforts are made by the countries which are developed to reduce the use of such fossil fuels but the same comes along with the costthat many countries could not afford to pay.

A carbon credit is authority to release a particular quantity of CO2 or other greenhouse gases. One credit allows someone to release One ton of CO2 or the equivalent in other greenhouse gases. The carbon credit is one side of a "cap-and- trade" mechanism. Carbon emitting companies have provided credits to continue polluting up to a certain point.

This restriction is adjusted on a regular basis. Meanwhile, the corporation may sell any credits that are no longer required toanother business that requires them. As aresult, private companies are motivated twice as much to minimize greenhouse gas emissions. If their emissions cross a certain limit, they should first pay money for more credits and subsequently make money by minimizing their emissions and selling the extra credits they possess.

II. AIMS AND OBJECTIVE

a) Aim

The purpose of this project is to create a model which can successfully predict CO2 emission based on varied input datasets with the analysis between two models at the least amount of cost possible and also to get high proficiency.

b) Objective

1.) This project's crucial intention is to provide the best Comparative analysis with the greatest degree of precession and the lowest cost function.
2.) To utilize minimum computational power.
3.) In a minimum amount of data to achieve higher efficiency.
4.) True challenge to build a system that shows the accurate level of prediction for CO2 emission from an environmental point of view.
5.) The output will be in numerical tableform and easy to analyze.

III. LITERATURE SURVEY

Paper1: Forecasting CO2 Emission with Machine Learning

This paper intended to provide information about atmospheric CO2 levels and how major greenhouse gas emissions, particularly CO2, influence the Earth, and how it can forecast this emission using ML (Machine

DOI : 10.35291/2454-9150.2022.0168

Learning). It's also important for figuring out how to anticipate CO2 emissions. The influence of greenhouse gases on global warming can therefore be mitigated. A gaseous layer envelops the earth and creates climatic changes in the atmosphere. The atmosphere comprises around 78% Nitrogen(N), 21% Oxygen(O2), 0.9% Argon (Ar), 0.03% Carbon-dioxide (CO2), and a trace quantity of noble gases like krypton, xenon, and helium. The greenhouse gas is generated by atmospheric gases that warm the planet.

Paper2: Carbon-dioxide emission prediction using the Support Vector Model

The SVM classifier was suggested in the literature to calculate the expense of Carbon- dioxide (CO2) outputs. The power consumption variable, which significantly affects the growth of CO2 emissions, were utilized to develop the model. They intendedto track CO2 emissions depending on thepower consumption used throughout the production process. The proportion of electricity utilized in training and testing the model was derived from the Alcohol Industry. It separated the training statistics into 90% and the testing statistic into 10%by using a cross-validation approach. In addition, the trial-and-error strategy in the experiment was utilized to discover the ideal parameters of the SVM model by modifying C parameters and Epsilon. In reality, thisresearch aided the business operator indeciding on the expenditure discharge of carbon dioxide.

Paper 3: Prediction Model for Carbon Dioxide Emissions and the Atmosphere

The contemporary study mission is to design a statistical model to calculate Carbon - dioxide emissions and also to study the condition of the atmosphere in the USA. They used monthly emissions information given by the CO2 Information Analysis Institute. It used statistics from the Scripps Institution in San Diego, which was acquired in Mauna Loa from 1965 to 2004to determine the amount of Co2 present in the atmosphere. Patterns and cyclical effectswere taken into seen in the statistical model that has been constructed. The trustability ofthe forecasting method is shown using real- world statistics.

IV. EXISTING SYSTEM

Support Vector Machine model was applied previously to predict Carbon-dioxide emissions from energy consumption. The model was used to track the quantity of Carbon-dioxide (CO2) released by power consumption and coal combustion. To get a better prediction model with a reduced error rate, a trial-and-error strategy was used. Prediction with high accuracy can give information concerning CO2 emissions. When creating the model, the major goal of constructing the new system is to get the lowest RMSE possible. It may be deduced that when the forecast model has a high precession, the lower RMSE value must be produced.

Limitations:

Higher RMSE with low accuracy.

SVM works better with distributed datasets, but accuracy decreases if a varying input set is provided.

IV. COMPARATIVE STUDY

Table 1 : Comparative Analysis

Sr. No.	Author	Paper Title	Publication	Purpose	Outcome
1.	Evin Garip and Ayse Betul Oktay	Forecasting CO2 Emissionwith MachineLearning	IEEE, 2018	This project aimed to predict the CO2 Emission in Turkey byusing machine learning methods.	The study produced analysis on svm an random forest for co2 prediction in which svm produced better result than random forest.
2.	Chairul Saleh, Nur Rachman Dzakiyullah and Jonathan BayuNugroho	Carbo dioxide emission prediction using the support vectormodel	IOP Publishing,2016	The goal of this researchis to track CO2emissions based on theelectrical energy andcoal utilized in the manufacturing process.	The collected study revealed that the least error value was 0.004 with optimal parameters for the SVM model of 0.1 for the C parameter and 0 for Epsilon.
3.	Shou Hsing Shih and Chris P. Tsokos	Prediction Models for Carbon Dioxide Emissions and the Atmosphere	International Journal, 2008	Using historical dataand the ARIMA method, the current work attempts to buildtwo distinct statistical models for carbonemissions and atmospheric carbondioxide in the USA.	The study outcome was to develop two non-stationary time series models with trend and seasonal effects to predict future estimates of carbon dioxide emissions and that in the atmosphere.

V. PROBLEM STATEMENT

The capacity to gather, organize, and to process data in the present system has proven tough. If Unstructured data is given this may lead arise to problems in the system. So, to remove the disadvantages in the existing system, this project is to overcome the inapplicability (irrelevancies) and improve the accuracy of this project.

1) If given a lot of datasets it increases load and sometimes even leads to model failure.

2) If the given dataset is not correct or not arranged properly or both then it will cause aproblem.

3) Multiple models cannot be run at a time so the efficiency decreases

4) Limited data has been tested constantly in this project.

VI. PROPOSED SYSTEM

The attribute-based input dataset is providedas initial input to the system. Post Successful data cleaning and normalization a cleaned input is provided to the Support Vector Machine based data regression model to predict the CO2 Emission forecast.Similar data is provided as input to the Random Forest for forecasting. The final result is compared with the Testing dataset to arrive at a better performance.

a) Time-based reference (Faster Algorithm based on varied input dataset)

b) Accuracy-based reference. (More accurate algorithm build on diverse input dataset)

VII. ALGORITHM

```
Step 1: Start
Step 2: Load csv file
path ← settings.MEDIA_ROOT + "//" +"owid-co2-data.csv"l ←
['co2_per_capita','coal_co2','coal_co2_per_capita','oil_co2','oil
_co2_per_capita','co2_growth_prct','co2']
Step 3: read csv file
df ← pd.read_csv(path)
df ← df[df.iso_code == "IND"]df ← df[l]
df ← df.fillna(0)
Step 4: data cleaning for dependent andindependent variables
X ← df.iloc[:, :-1].values          # indipendent variable
y ← df.iloc[:, -1].values           # Dependent variable
Step 5: Create training and testing samples X_train, X_test, y_train, y_test ←
train_test_split(X, y, test_size=0.2, random_state=0)
Step 6: Fit the model
Model←tree.DecisionTreeRegressor(random_state=0)
clf ← model.fit(X_train, y_train)
Step 7: Predict class labels on training data pred_labels_tr ←
model.predict(X_train) Step 8: Predict class labels on a test
data pred_labels_te ← model.predict(X_test) score_te ←
model.score(X_test , y_test) Step 9: Look at classification
report toevaluate the model
score_tr ← model.score(X_train, y_train)
Step 10: RF implementation model ←
RandomForestRegressor()y_pred ← model.predict(X_test)
mae ← .mean_absolute_error(y_test,y_pred)
mse ← .mean_squared_error(y_test, y_pred)r2_score ←
.r2_score(y_test, y_pred)
rmse ← sqrt(_mean_squared_error(y_test,y_pred))
rslt_dict ← {"mae": mae,
"mse": mse, "r2_score": r2_score,"rmse": rmse
}
Step 10: SVM implementationmodel ← SVR()
model.fin(X_train, y_train) y_pred ← model.predict(X_test)
mae ← .mean_absolute_error(y_test,y_pred)
mse ← _mean_squared_error(y_test, y_pred)r2_score ←
.r2_score(y_test, y_pred)
rmse ← sqrt(_mean_squared_error(y_test,y_pred))
rslt_dict ← {"mae": mae,
"mse": mse, "r2_score": r2_score,"rmse": rmse
}
Step 11: Forecasting
path ← settings.MEDIA_ROOT + "\\" +"owid-co2-data.csv"
self.df ← pd.read_csv(path)
self.df ← self.df[self.df.iso_code == "IND"]df ←
self.df[['year', 'co2']]
df['year'] ← pd.to_datetime(df['year'])
dp = pd.to_datetime(df['year'], format='%Y')df ←
df.groupby(dp)['co2'].sum().reset_index()df ←
df.set_index('year')
df.index
y ← df['co2'].resample('MS').mean()y['2018':]
p ← d ← q ← range(0, 2)
pdq ← list(itertools.product(p, d, q)) seasonal_pdm ←
[(x[0], x[1], x[2], 12) for xin list(itertools.product(p, d, m))]
for param in pdm:
for param_seasonal in seasonal_pdm:try:
mod ← sma.tsp.statespace.SARIMAX(y,order← param,
seasonal_order← param_seasonal,enforce__stationarity←
False, enforce__invertibility← False) results ← mod.fit()
except Exception as e:continue mod ←
sma.tsp.statesspace.SARIMAX(y,order← (1, 1, 1),
seasonal_order← (1, 1, 0, 12),enforce_stationarity← False,
enforce_invertibility← False)results ← mod.fin()
pred_uc ← results.get_forecast(steps=800)pred_ci ←
pred_uc.conf_int()
Step 11: Displaying the result
Step 12: End
```

VIII. MATHEMATICAL MODEL

Support Vector Machine (SVM):

Lately, numerous SVM implementations for regression and classification work have beendiscovered. SVM is a advance machine learning technique that can be used to forecast time series.

The SVM algorithm's purpose is to find the ideal line or decision boundary for categorizing n-dimensional space so that it may simply place fresh data points in the right classification in the time ahead. A hyperplane is the optimal

DOI : 10.35291/2454-9150.2022.0168

choice boundary.

SVM chooses the greatest points/vectors which aid in the creation of the hyperplane. These extreme examples are also regarded as support vectors, and the technique is known as the Support Vector Machine.

In machine learning model, deciding the model's accuracy is a vital step. To calulate the model's performance in regression analysis, the Mean Absolute Error, Mean Square Error, Rooted Mean Square Error and determination coefficient metrics are utilized.

The Mean absolute error is the arithmetic mean of the difference between the actual and forecasted values in the dataset. It calculates the mean of the dataset's remnant.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|xi - x|$$

Where , xi – predicted value

x – mean value of x

[MSE]Mean squared error is defined as the averaged squared difference between the current and anticipated values of the given dataset (mse). It calculates the residuals' variance.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{x_i} - \mathrm{x})\,2$$

The square root of Root Mean Squared Error is the square root of Mean Squared Error. It calculates the residuals' standard deviation.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{x_i} - \mathrm{x})2}\;.$$

The coefficient of determination, commonly referred as R-squared, is the proportion of variance of the dependent variable that can be explained by a linear regression model. It is a scale-free score, meaning that regardless matter how little or huge the values are, R square will be less than one.

$$\mathrm{R}^2 = 1 - \frac{\sum(\mathrm{x_i} - \mathrm{x})^2}{\sum(\mathrm{x_i} - \mathrm{x})^2}$$

Modified R squared is a modified form of R square that takes into account the number of independent variables in the model and is always equal to or less than R2. The amount of observations in the data is n, and the amount of independent variables is j in the below equation

$$R^2_{adj} = \left[\frac{(n - R^2)(n - 1)}{n - j - 1}\right]$$

Random forest (RF):

Random forest is a supervised classification approach that builds a forest while also turning it randomized in some way. The more trees there are, the more accurate the outcome, which may be utilized in both classifications and also for regression. The random forest classifier can manage missing data and models for categorical values.

Random Forest operates in two phases:

1)The starting step is to create a Random Forest.

a.) Pick K features at random amongst m features.

b.) The node is broken into daughter nodes.

c.) Repeat steps a–c till the amount of nodes obtained is one.

d.) By repeating the procedure from a through a number of times to produce a number of trees. As a result, a forest grows.

2. The following stage would be to make a forecast using a random forest classifier.

A result is anticipated and saved by utilizing testing features as well as the norms of each randomly generated decision tree. For each predicted goal, votes are calculated.

IX. SYSTEM ARCHITECTURE

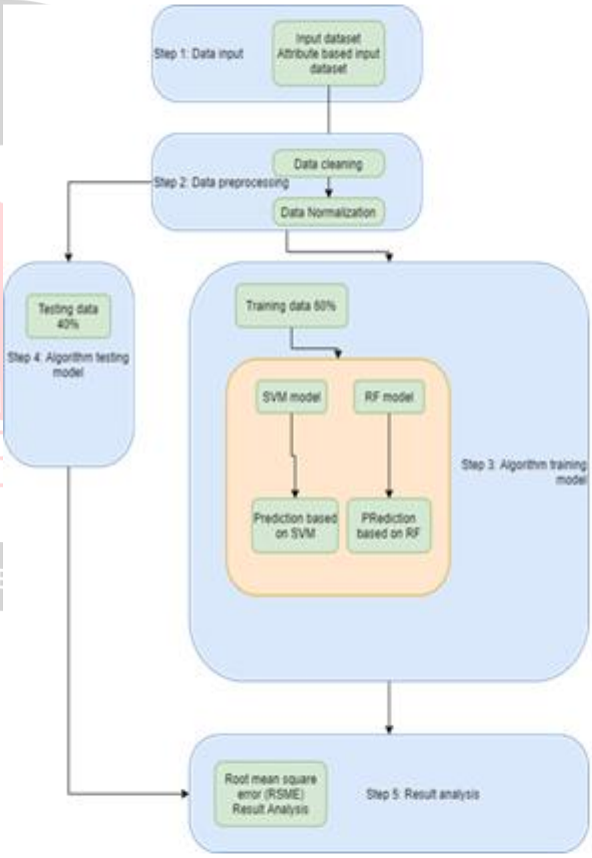

Step 1: Data input
Input dataset Attribute based input dataset
Data cleaning
Step 2: Data preprocessing
Data Normalization
Testing data 40%
Step 4: Algorithm testing model
Training data 60%
SVM model
RF model
Prediction based on SVM
PRediction based on RF
Step 3: Algorithm training model
Root mean square error (RSME) Result Analysis
Step 5: Result analysis

Fig.1: System Architecture

Description:

Data Input: The Input dataset is provided to the models in the form of a .csv file.

Data Processing: In the .csv file if there are any corrupted, incorrectly formatted, duplicated, or incomplete datasets then it is removed and then this dataset is organized to look similar across all fields.

DOI : 10.35291/2454-9150.2022.0168

Algorithm Training: In this process of training an ML model involves providing an ML algorithm (that is, Random Forest and Support Vector Machine) with 60% training data to learn from. The learning algorithm finds patterns in the training data that map the input data attributes to the target (the answer that you want to predict), and it outputs an ML model that captures these patterns.

Algorithm Testing: It is the actual execution of the algorithm. Where the User provides input (that is, 40% of the dataset) and the algorithm calculates the output based on the learned parameters from the training phase.

Result Analysis: In this step, the output is analyzed on the bases of MAE, MSE, RMSE, and R2- scores.

X.ADVANTAGES

1.) Its helps to solve complicated real-world issues with several constraints.

2.) It reduces time in calculation and also reduce any unwanted effort.

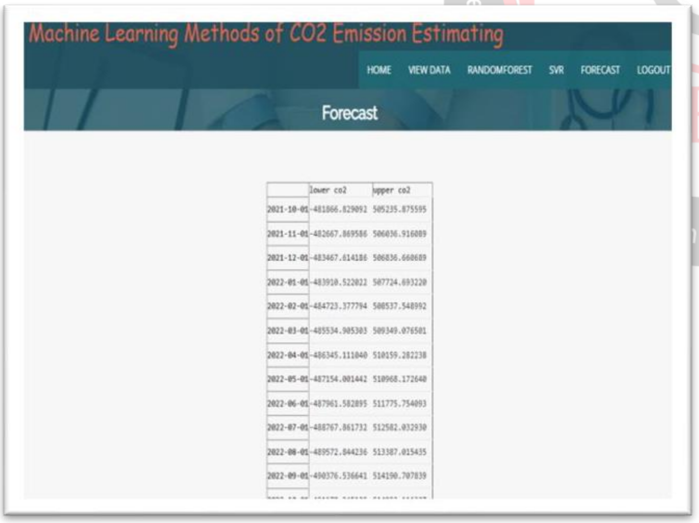
3.) Its gives a fast and accurate prediction.

4.) It recognizes the crucial characteristics without user intervention.

5.) Its uses less computational power or resources like RAM, CPU, GPU, TPU, etc

6.) Given less quantity of data, It can achieve more accuracy.

XII. DESIGN DETAILS



	lower co2	upper co2
2021-10-01	-481866.829091	505235.875595
2021-11-01	-482667.869586	506036.916089
2021-12-01	-483467.614186	506836.660689
2022-01-01	-483910.522021	507724.693220
2022-02-01	-484723.377794	508537.548992
2022-03-01	-485534.905303	509349.076501
2022-04-01	-486345.111040	510159.282238
2022-05-01	-487154.001442	510968.172640
2022-06-01	-487961.582895	511775.754093
2022-07-01	-488767.861732	512582.032930
2022-08-01	-489572.844236	513387.015435
2022-09-01	-490376.536641	514190.707839

Fig.2: Forecast Result

Description:

This project website is based on Django and Python. In the above, Fig.2 Forecast Page is shown where all the predicted values of CO2 emission are shown ranging from lower CO2 value to upper CO2.

XIII. CONCLUSION

Thus we have tried to implement the paper "Chairul Saleh1, Nur Rachman Dzakiyullah2 and Jonathan Bayu Nugroho Fellow", "Carbon dioxide emission prediction using support vector machine ", IOP(2016) and according to implementation, the conclusion is to successfully forecast the co2 discharge. In this paper, the ML model has performed the support vector machine and random forest algorithm to forecast the carbon dioxide discharge. It also tells which algorithm (that is, random forest and SVM) is best for CO2 emission prediction. The model is further compared to show the best performance to forecast based on RMSE, MAE, MSE, and R2 parameters. Hence the above project implemented is basically to see which model is best for the forecast of CO2 emission.

REFERENCE

[1] Chairul Saleh, Nur Rachman Dzakiyullah and Jonathan Bayu Nugroho ," Carbo dioxide emission prediction using the support vector model ", IOP Publishing, 2016

[2] Evin Garip and Ayse Betul Oktay,"Forecasting CO2 Emission with Machine Learning ", IEEE, 2018

[3] CShou Hsing Shih and Chris P. Tsokos ," Prediction Models for Carbon Dioxide Emissions and the Atmosphere ", International Journal, 2008

[4] Lakshay Amarpuri, Navdeep Yadav, Girish Kumar and Saurabh Agrawal," Prediction of CO2 emissions using deep learning hybrid approach", IEEE, 2019

[5] Suhasini Vijayumarand Pooja Kadam," Prediction Model: CO2 Emission Using Machine Learning", Research Gate, 2018

[6] S. Yeasmin, S. N. J. Syed, L. A. Shmais and R. A. Dubayyan, "Artificial Intelligence-based CO2 Emission Predictive Analysis System", International Conference on Artificial Intelligence & Modern Assistive Technology (ICAIMAT), 2020

[7] L. Amarpuri, N. Yadav, G. Kumar and S. Agrawal, "Prediction of CO2 emissions using deep learning hybrid approach: A Case Study in Indian Context",Twelfth International Conference on Contemporary Computing (IC3), 2019

[8] T. C. Ho, S. C. K. M. Z. M. Jafri and L. H. San, "A prediction model for CO2 emission from the manufacturing industry and construction in Malaysia", International Conference on Space Science and Communication (IconSpace), 2015

[9] S. Kangralkar and R. Khanai, "Machine Learning Application for Automotive Emission Prediction," 6th International Conference for Convergence in Technology (I2CT), 2021.

[10] T. C. Ho, S. C. Keat, M. Z. M. Jafri and L. H. San, "A prediction model for CO2 emission from manufacturing industry and construction in Malaysia", Int. Conf. Sp. Sci. Commun. Iconsp, September 2015

DOI : 10.35291/2454-9150.2022.0168