 ISSN NO : 1006-8341

Analysis of Carbon Emissions in Food System using Machine Learning

Dr. Bharati Wukkadada
K J Somaiya Institute of Management, Mumbai
Nivedita Bhatia
K J Somaiya Institute of Management, Mumbai

ABSTRACT

Global climate change is mostly caused by carbon dioxide emissions. It is commonly acknowledged that, in order to avert the worst effects of climate change, the world must reduce emissions as soon as possible. However, how this obligation is shared across regions, countries, and individuals has long been a source of debate in international forums. Carbon emissions is the amount of carbon element released from human activity. The greater the emission, the more significant impact it has on the environment. The paper presents the method to identify whether the countries are developed or not based on their carbon emissions/footprint throughout the years. This evaluation is essentialized machine learning methods, i.e., Decision trees, Logistic Regression, and KNN. Through these methods, we can classify the developed or industrialized countries based on the substances produced by the food system. The data used is from EDGAR- global emission inventory for food systems. This research demonstrates the impact of countries collectively on the environment from the food sector.

Keywords

Carbon, Carbon emissions, food, greenhouse gas, carbon footprint, classification, logistic regression, Decision tree, edgar.

1. INTRODUCTION

Food is the necessity to survive and grow on this planet earth. Even the basic need of survival is incorporated of carbon. Everything in this world is compounded with carbon element to maintain the balance the life on the planet earth. Not only driving vehicles or producing smoke, even our lifestyle choices, the food we eat impacts our overall carbon footprint. If we talk about the global emissions, one quarter of it is, is caused by food production and agriculture itself.

Carbon footprint is the amalgamation of greenhouse gas emissions caused by any form of living on this earth. The natural carbon emissions are caused by soil, ocean, etc which supports life on earth. But when humans extract, refine, transport or burn fossils we release an extra amount or greenhouse gas with cutting a source absorbing CO2 leading to greater catastrophic effects. Carbon emissions can be controlled at the consumer level, as producing demand drives production. The increase in greenhouse gas emissions is increase in human activities which results in greater carbon footprint of an individual, organization or a community. The increase in carbon footprint gives us the climate change, droughts, melting glaciers, heatwaves etc.

2. LITERATURE REVIEW

The concept of carbon footprint was developed in the 1990s. It can be measured through various carbon accounting denoted assessments. Food system is responsible for one-third of global greenhouse emission. Several research have been done since the concept came into light. The development of strategies to reduce the carbon footprint of the food system was one of the research programmes in China. It argued that technical means alone will not be able to significantly reverse the direction of climate change, as consumer behaviour has a significant impact on climate change. In other one of the research projects in Spain depicted how dietary choices or sustainable diets can differentiate in environmental sustainability. Since carbon emission with the global atmosphere also effects the health and well-being of humans, several footprint calculating software are also available with increased risk over decade.

3. METHODOLOGY

EDGAR is the global database of human induced emissions of greenhouse gases. The food system globally comprehends necessary economic sectors making sufficiently great contribution to greenhouse emissions. The food system emits far more greenhouse gases than land-based system. Food must be grown, harvested, or captured, transported, processed, packed, distributed, and cooked, with any leftovers disposed of properly. The entire process, from production to disposal, necessitates a tremendous amount of energy. This energy must be available at any given time and in any given location.

```
df.head()
```

	Country_code_A3	Name	C_group_IM24_sh	dev_country	Substance	FOOD_system_stage	Y_1990	Y_1991	
0	AGO	Angola	10:_Southern_Africa	D	GWP_100_CO2	LULUC (Production)	10273.5451	10273.5451	10
1	AGO	Angola	10:_Southern_Africa	D	GWP_100_CH4	LULUC (Production)	7662.4856	7662.4856	7
2	AGO	Angola	10:_Southern_Africa	D	GWP_100_N2O	LULUC (Production)	5009.7985	5009.7985	5
3	ALB	Albania	12:_Central Europe	I	GWP_100_CO2	LULUC (Production)	116.8068	116.8068	
4	ALB	Albania	12:_Central Europe	I	GWP_100_N2O	LULUC (Production)	12.4815	12.4815	

5 rows × 35 columns

Fig 1. Data Exploration

Using a collated database of food system from 1990-2018, countries were classified into two categories; Developed or Industrialized. Four emission substances were found in eight stages of food production. The substances which were contributing to the emission are CO2, CH4, N2O and F-gases. Data of each stage of production was recorded on yearly basis. The predictive models were created using Logistic Regression, Decision Tree Classifier and K-nearest neighbours to identify the country class. In order to design the prediction model data was pre-processed to improve the performance of the model. The pre-processing of data is normalizing raw data or recorded data in order to make it more understandable. The data was divided into training and testing sets. Training Data comprises of 90% of features of data set whereas testing data consists 10% of target of dataset to validate the model. Cross-Validation technique was used for the validation.

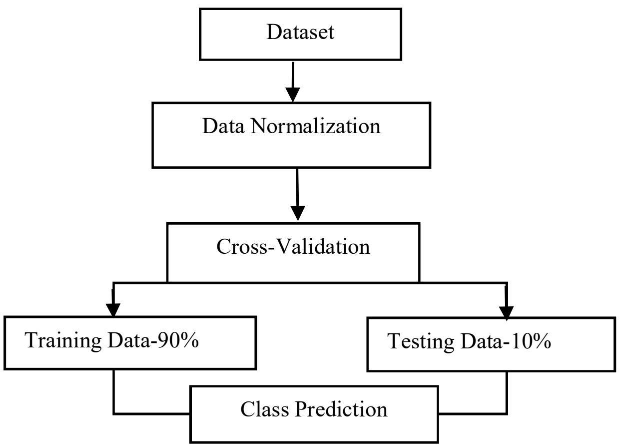

Fig 2. Data Flow

The model was designed to find how the emissions evolves through the different stages of food production system with the growing population and dietary changes and needs. The prediction of country class depends on the evolvement of these various factors. Four features having an object value were label encoded using the label encoder, before splitting the data. Three classification models were used to find the model with the better performance regardless of same data. The three of the models were logistic regression, decision tree and KNN. The logistic regression model is used to predict the binary outcome which is either yes or no. Whereas, the decision tree model classifies the binary outcome by deciding it from the root and leaf nodes. In the KNN model the binary outcome is determined by calculating the distance between test and training set with identifying the K no of points nearest to the test data. The model with the highest precision was decision tree with accuracy rate of 95% iterated with two max leaf nodes. KNN with 94% accuracy rate was iterated with five nearest neighbors. The model with least accuracy rate was logistic regression with 73% accuracy. The data was scaled using the Standard scaler before being trained in the model.

4. RESULTS

The objective was to attain the model with highest accuracy classifying the country class on dependency of data obtained from lifecycle of food system. It was obtained using the classifier model, decision tree, which resulted in highest precision and accuracy rate. The slight shift was observed between two decades of data. The image below (Fig 3) shows the classification report of Logistic Regression with the accuracy rate of 73%.

```
[[704   2]
 [259   2]]
              precision    recall  f1-score   support

           0       0.73      1.00      0.84       706
           1       0.50      0.01      0.02       261

    accuracy                           0.73       967
   macro avg       0.62      0.50      0.43       967
weighted avg       0.67      0.73      0.62       967

73.00930713547052
```

Fig. 3. Classification Report for Logistic Regression

The below image (Fig. 4.) shows the classification report of Decision tree with the accuracy rate of 95%.

```
[[706   0]
 [ 43 218]]
              precision    recall  f1-score   support

           0       0.94      1.00      0.97       706
           1       1.00      0.84      0.91       261

    accuracy                           0.96       967
   macro avg       0.97      0.92      0.94       967
weighted avg       0.96      0.96      0.95       967

95.55325749741469
```

Fig. 4. Classification Report for Decision Tree

The below image (Fig. 5.) shows the classification report of K-Neighbors with the accuracy rate of 94%.

```
[[699   7]
 [ 43 218]]
              precision    recall  f1-score   support

           0       0.94      0.99      0.97       706
           1       0.97      0.84      0.90       261

    accuracy                           0.95       967
   macro avg       0.96      0.91      0.93       967
weighted avg       0.95      0.95      0.95       967

94.82936918304034
```

Fig.5. Classification Report for K-NeigborsClassifier

5. CONCLUSION

Greenhouse gas emissions have a significant influence on the environment. The global growth in carbon emissions from numerous businesses has an impact on every element. With the recorded EDGAR data of greenhouse emissions at various phases of the food manufacturing system, the overall model aids in identifying the country's class. Adding EDGAR-FOOD to emission-reduction strategies is critical for anticipating future changes in the whole food system and building effective mitigation techniques to prevent non-targeted emissions.

6. REFERENCES

[1] Carbon emission of food consumption: An empirical analysis of China's residents

[2] Application of Machine Learning Methods in Carbon Footprint Optimization

[3] Machine-Learning-Based Carbon Footprint Management in the Frozen Vegetable Processing Industry