**PREDICTION OF CO2 EMISSIONS BY COUNTRY USING IBM WASTON**

**1. INTRODUCTION**

1.1 OVERVIEW

Growth comes in tandem with industrialization, which involves the use of energy, leading to increased carbon emissions and environmental degradation. Therefore, as a country industrializes, its pollution levels will increase significantly. This is the natural relationship between economic activity and the environment. Less developed countries pollute less because they generally have low incomes, so people are unable to afford products which contribute to increased pollution, such as cars, planes, televisions, etc. However, when people improve their quality of life, they are able to buy more luxury goods and services, leading to increased pollution. The Environmental Kuznets Curve hypothesis combines this trend and the idea that post-industrialized countries invest and use alternative energy sources. The Kuznets Curve hypothesis states that as countries industrialize, they emit larger amounts of carbon dioxide. But once countries become more advanced industrially, they look to cleaner sources of energy, reducing pollution. Also, un-industrialized countries do not pollute as much since they do not consume much energy. In a previous study, I analysed the relationship of GDP on Carbon Emissions (the Environmental Kuznets Curve) for the years 1994, 2004, and 2014. I saw a negative-quadratic relationship between GDP and Carbon Emissions, matching the trend established in the Environmental Kuznets Curve hypothesis.

In this paper, I use Machine Learning to predict CO2 emission of a country using 9 years of data. The model uses a handful of variables to predict the CO2 emissions of a country. The purpose of this model is to make accurate CO2 emission predictions given a few variables, using the Kuznets Curve as the foundation.

1.2 PURPOSE

Carbon emission research was initiated in 1981 with the first publication focusing on volatile organic carbon emissions of cooling tower water [22]. Thereafter several researchers explored the domain without significant impacts until the Kyoto protocol was signed in 1997. However, carbon emission research became a trending topic after 2007, resulting in a significant increase in research publications. With the rising global temperatures, climate change has become a major global concern which is also considered as the most serious issue global community has to address in the 21st century [23]. Research on global carbon emissions has significantly increased after discovering carbon emissions as the major cause of climate change.

Accordingly, it is evident that the majority of carbon emission research relates to environment related aspects. Economics, energy fuels, ecology, and civil engineering are some of the other notable research areas which have been affected by carbon emission research. A major reason for carbon emission research to expand into the aforementioned research areas is the goal of sustainable development which has become increasingly popular among the global community due to low carbon emission society concept.

**2. LITERATURE SURVEY**

2.1 EXISTING PROBLEM

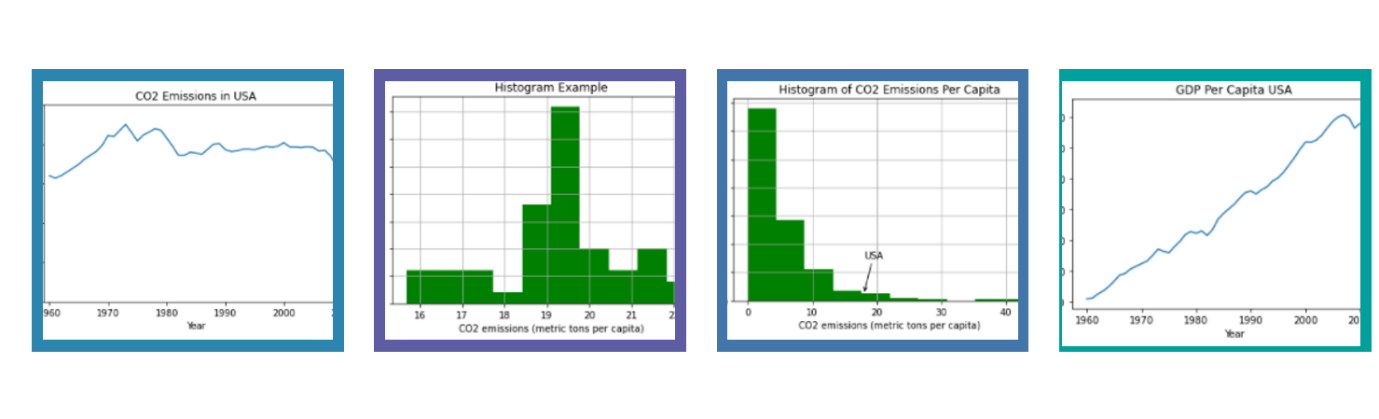
Economic scale, economic structure, and technological level are the three major factors affecting the environment Economic scale is the output of the economy; more economic output means more pollution. It is because that economic growth needs more resources investment and more energy consumption. Economic structure is industry structure. The change of industry structure will reduce the pollution. With economic developing, percentage of secondary industry, especially energy-intensive industry, will reduce percentage of tertiary industry, and energy consumption will increase, so the pollution will be reducing. Technology progress will realize the usage of resource efficiency and reduce the energy consumption. So, technological level is an important factor which influences the energy intensity and pollution. Many research studies are based on these three environment factors and extend them accordingly. The research can be classified into relation between economic growth and CO2 emissions, industry structure, technology, and CO2 emissions, and urbanization and CO2 emissions. But results differ from research focus, theories, and methods. There are three parallel literatures on factors what will influence CO2 emissions.

2.2 PROPOSED SOLUTION

Logistic regression

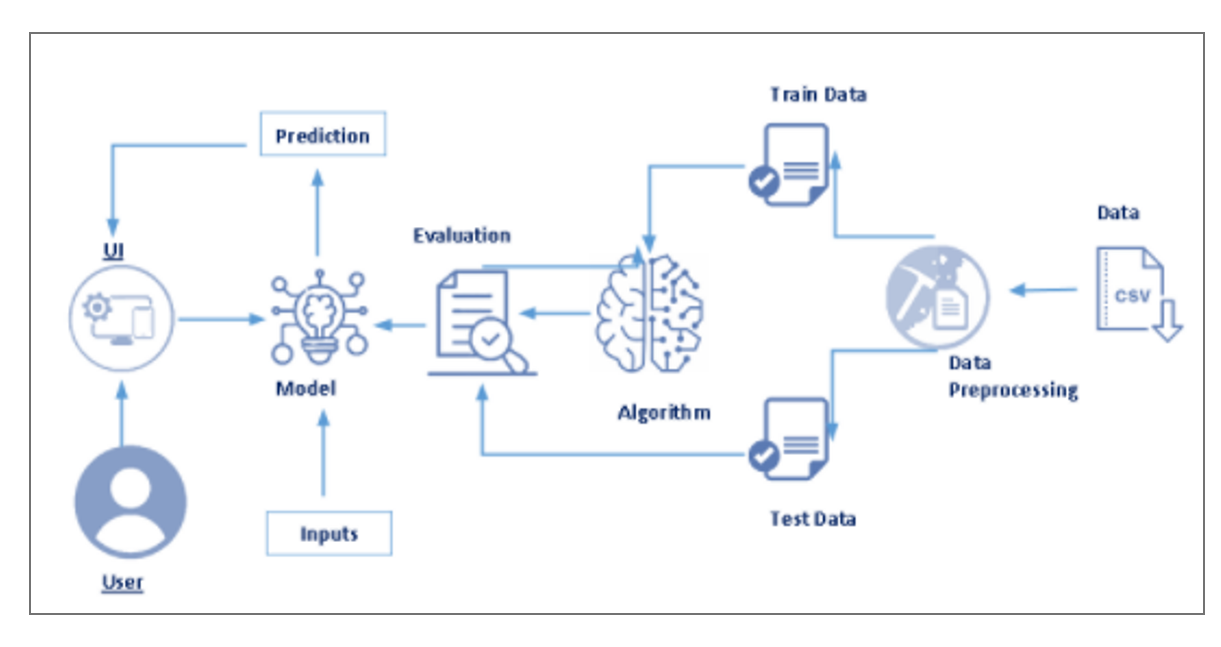
The logistic regression (LR) algorithm is used for supervised learning, and widely used for binary classification tasks. It is a branch of natural language processing (NLP), which is generally thought of as part of artificial intelligence (AL). LR permits obtaining insights about the model, such as observed coefficients Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a dataset. Logistic regression has become an important tool in the discipline of machine learning. The approach allows an algorithm being used in a machine learning application to classify incoming data based on historical data. As more relevant data comes in, the algorithm should get better at predicting classification within data sets. Logistic regression can also play a role in data preparation activities by allowing data sets to be put into specially predefined buckets during the extract, transform, load (ETL) process in order to stage the information for analysis. A logistic regression models predicts a dependent data variable by analysing the relationship between one or more existing independent variables. For example, a logistic regression could be used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted to a particular college.

After observing storing linear relationships between economic variables and Transportation emissi0ns, bivariate linear regression relationships were formally defined. These regressions used 2016 data, and again removed the country outliers that must be studied individually. Python scripts were connected to the .csv tables and used to build the regression models. Python packages pandas, pyplot, matplotlib, and scikit learn were used for the model and visualization.

**3. THEORITICAL ANALYSIS**

3.1 BLOCK DIAGRAM

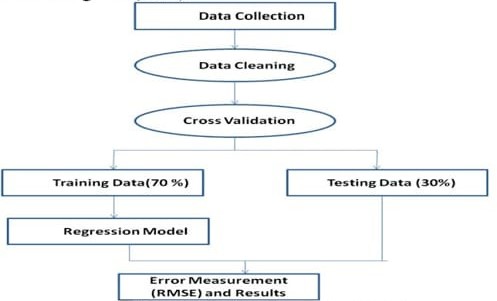
**ARCHITECTURE**

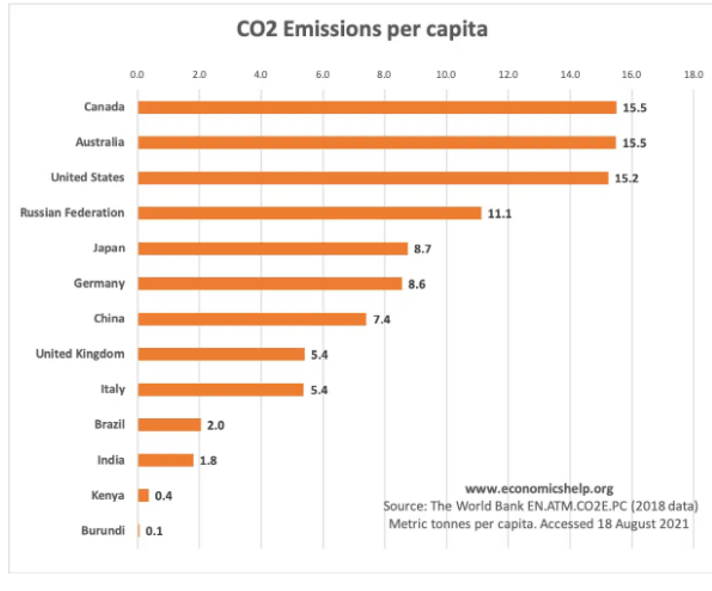


**4. EXPERIMENTAL INVESTIGATIONS**

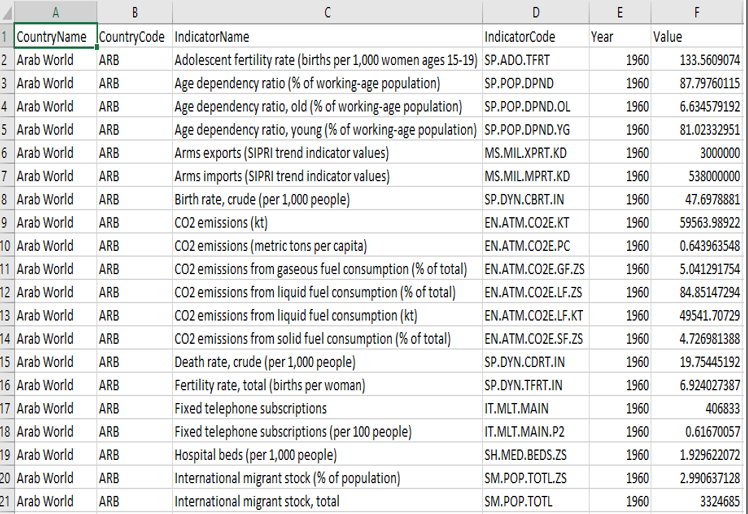
The experimental analyses shown that choosing appropriate combination may result in significant improvement on classification accuracy. NOBATA& Tetreault used normalization of numbers, replacing very long unknow words and repeated punctuations with the same token. Explained the role of transformation in sentiment analyses and demonstrated with the help of SVM on movie review database that the accuracies improve significantly with the appropriate transformation and feature selection. They used transformation methods such as white space removal, expanding abbreviation, stemming, stop words removal and negation handling. Other works focus more on modelling as compared to transformation. In the study, Bojanowski et al. used five transformations namely URLs features reservations, negation transformation, repeated letter normalization, stemming and lemmatization on twitter data and applied liner classifier available in WEKA machine learning tool. They found the accuracy of the classification increases when URLs features reservations, negation transformation and repeated letters normalization are employed while decreases when stemming and lemmatization are applied. The investigation the effect of transformation on five different twitter datasets in order to perform sentiment classification and found that removal of URLs, the removal of stop words and the removal of numbers have minimal effect on accuracy whereas replacing negation and expanding acronyms can improve the accuracy. Most of the expanding regarding application of the transformation has been around the sentimental classification on twitter data which is length restricted.

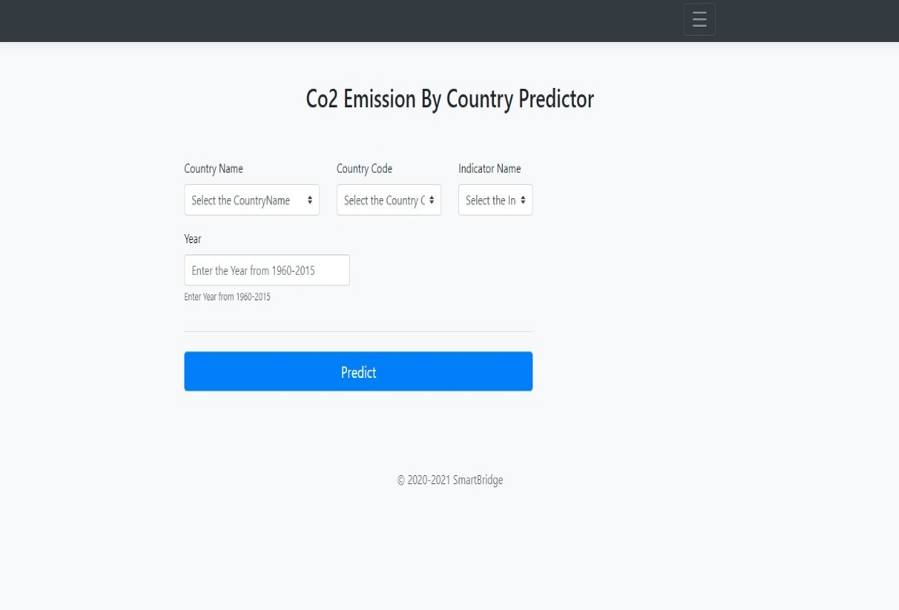
**5. FLOWCHART**





**6. RESULT**





Here we used logistic regression to work out on a predicting and analysing co2 emissions by country.

After giving the suitable inputs we get the output as shown below.



**7. ADVANTAGES AND DISADVANTAGES**

ADVANTAGES BASED ON PROPOSED SOLUTIONS

Technology more mature than other alternatives; can easily retrofit into existing plants; High co2 concentration enhance sorption efficiency; fully developed technology, commercially deployed at the required scale in some industrial sectors; opportunity for retrofit to existing plant.

Very high co2 concentration that enhances absorption efficiency; mature air separation Technologies available; hence required smaller boiler and other equipment; co2 is the main combustion product, which remains unmixed with N2, thus avoiding energy intensive air separation.

Here's the big, important thing about CO2: It's a greenhouse gas. That means CO2 in the atmosphere works to trap heat close to Earth. It helps Earth to hold on to some of the energy it gets from the Sun so the energy doesn't all leak back out into space.

DISADVANTAGES

Entry costs still high, although those of wind have fallen fast and the efficiency of solar cells has improved; climate and geography may impose a niche role; large hydro projects contested because of environmental impact.

The amount of carbon emissions trapped in our atmosphere causes global warming, which causes climate change, symptoms of which include melting of the polar ice caps, the rising of sea levels, the disturbance of animals' natural habitats, extreme weather events, and so many more negative side effects that are dangerous

These carbon emissions raise global temperatures by trapping solar energy in the atmosphere. This alters water supplies and weather patterns, changes the growing season for food crops and threatens coastal communities with increasing sea levels.

**8. APPLICATIONS**

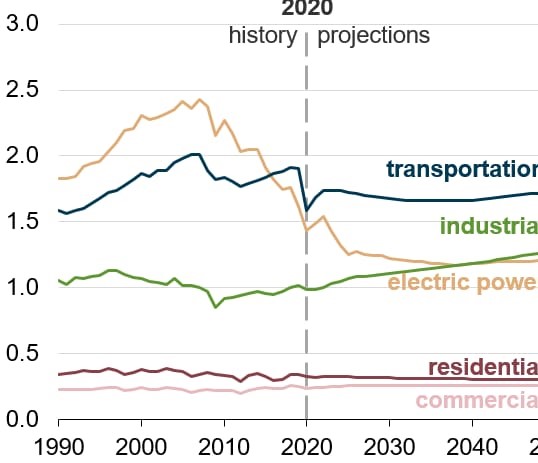
Carbon dioxide is used as a refrigerant, in fire extinguishers, for inflating life rafts and life jackets, blasting coal, foaming rubber and plastics, promoting the growth of plants in greenhouses, immobilizing animals before slaughter, and in carbonated beverages.

**9. CONCLUSION**

This can be applied to predict CO2 emissions from energy consumption which can give us more accuracy. This model is used to monitor electrical energy and burning coal which affect the amount of CO2 emitted. It can be concluded that with the high accuracy of the prediction model, then the lower RMSE value must be obtained. For better results we used random forest regressor algorithm.

**10. FUTURE SCOPE**

After 2035, U.S. CO2 emissions begin to trend upward, reflecting the overall increase in the use of energy as a result of increasing population and economic growth. EIA projects that total U.S. energy-related CO2 emissions in 2050 will be about 4,807 million metric tons, or about 5% more than the amount in 2020.



**11. BIBILOGRAPHY**

1. L. Shipper, L. Scholl, and L. Price, “Energy use and carbon emissions from freight in 10 industrialized countries: an analysis of trends from 1973 to 1992,” Transportation Research Part D: Transport and Environment, vol. 2, pp. 57–76, 1997.View at: Google Scholar.

2. G. R. Timilsina and A. Shrestha, “Transport sector CO2 emissions growth in Asia: underlying factors and policy options,” Energy Policy, vol. 37, no. 11, pp. 4523–4539, 2009. View at: Publisher Site Google Scholar.

3. G. R. Timilsina and A. Shrestha, “Factors affecting transport sector CO2 emissions growth in Latin American and Caribbean countries: an LMDI decomposition analysis,” International Journal of Energy Research, vol. 33, no. 4, pp. 396–414, 2009.View at: Publisher Site | Google Scholar.

4. B. Talbi, “CO2 emissions reduction in road transport sector in Tunisia,” Renewable and Sustainable Energy Reviews, vol. 69, pp. 232–238, 2017. View at: Publisher Site | Google Scholar.

5. C. Zhu and D. Gao, “A research on the factors influencing carbon emission of transportation industry in “the belt and road initiative” countries based on Panel data,” Energies, vol. 12, no. 12, p. 2405, 2019.View at: Publisher Site | Google Scholar.

6. Y. Liang, D. Niu, H. Wang, and Y. Li, “Factors affecting transportation sector CO2 emissions growth in China: an LMDI decomposition analysis,” Sustainability, vol. 9, no. 10, p. 1730, 2017. View at: Publisher Site | Google Scholar.

7. Y. Wang, Y. Zhou, L. Zhu, F. Zhang, and Y. Zhang, “Influencing factors and decoupling elasticity of China’s transportation carbon emissions,” Energies, vol. 11, no. 5, p. 1157, 2018.View at: Publisher Site | Google Scholar.

**APPENDIX**

1. Source code

https://www.kaggle.com/ashukr/exploring-co2-emission?select=Indicators.csv