PREDICTION OF CO2 EMISSION LEVELS IN INDIA USING MACHINE LEARNING TECHNIQUES

Aishitha Pachipala (2120425)

*Authors*

This study attempts to examine the dynamics of CO2 emissions in India over a span of 25 years over the period 1998–2022, emphasizing the interplay between urbanization, energy consumption, and trade openness. Employing the Autoregressive Distributed Lag (ARDL) approach, we analyze the long-term relationship and short-term dynamics among these variables and CO2 emissions. After understanding the factors affecting CO2 emissions in the country we deploy supervised machine learning models such as Decision Tree and Support Vector Machine to forecast CO2 emissions, to compare prediction accuracy of the machine learning models and to select the best model based on the highest prediction accuracy along with offering insights into future trajectories.

*Introduction*

• To understand various factors that is driving carbon emissions in India and analyzes their impact.

• To predict CO2 emission levels using supervised machine learning techniques and comparing their prediction accuracy to select the best model.

*Objective*

We utilize the ARDL approach to examine the dynamics of the relationship between CO2 emissions and other variables. First, we conduct a unit root test to ensure stationarity, followed by applying an ARDL bound test to determine cointegration among the variables. We then deploy an ARDL model for analysis. Furthermore, we train the data using both decision tree model and support vector machine model. After running the models, we evaluate the accuracy of the predictions through testing.

*Methodology*

Among the independent variables, only fossil fuel energy consumption (Ln Ec) demonstrated a statistically significant impact on CO2 emissions, as indicated by the smaller p-value associated with its coefficient. Both supervised machine learning models demonstrated effective performance in predicting CO2 emissions. But however, based on the comparison, the Decision Tree Classifier appeared to be slightly superior in terms of accuracy and closeness to actual values

*Results*

The variables are found to be in first difference order, indicating that they are not stationary and lie in I(1). The ARDL Bound Test results suggest that the variables are not likely cointegrated at the specified significance level. The ARDL regression model shows that fossil fuel energy consumption has a significant impact on CO2 emissions. Additionally, the diagnostic tests, including autocorrelation and heteroscedasticity, suggest moderate autocorrelation and no heteroscedasticity in the data. Lastly, the Decision Tree Classifier model effectively predicted CO2 emission values for various years, with a high accuracy of 96.06% and the comparison between actual and predicted CO2 emission values suggested that the model fits the data well. While, Support Vector Machine model performed well, with low error metrics and a high r2 score of 91.55%, indicating a significant proportion of the variance. But however, the SVM model's predictions showed slightly higher errors (10.58%) compared to the Decision Tree Classifier’s predictions (7.23%). Based on the comparison, the Decision Tree Classifier appeared to be slightly superior in terms of accuracy and closeness to actual values.

*Analysis*

Overall, the findings suggest that reducing the energy consumption of fossil fuels can contribute to the reduction of carbon dioxide emissions in the study country. In conclusion, both the Decision Tree Classifier and the Support Vector Machine models demonstrated effective performance in predicting CO2 emissions. But however, based on the comparison, the Decision Tree Classifier appeared to be slightly superior in terms of accuracy and closeness to actual values.

*Conclusion*