2024

CAR CO2 EMISSIONS

USING TIME SERIES ANALYSIS

CREATED BY:

- *Ahmed Yasser Abdallaha El Sharkawy*
- *Mohamed Nasr Nasr*
- *Mohamed Mahmoud Fathallah*

INTRODUCTION

In this project, we aimed to analyze and predict carbon dioxide (CO2) emissions from cars worldwide using time series analysis. By understanding the trends and making accurate predictions, we can better comprehend the impact of automotive emissions on the environment and potentially inform policy decisions to reduce these emissions. We used a dataset containing CO2 emissions data, processed it, and applied various regression models to predict future emissions.

METHODOLOGY

DATA LOADING AND PREPROCESSING

We started by loading the dataset using pandas and then normalized the value of Co2 emissions column using MinMaxScaler to ensure the values are scaled between 0 and 1.

We plotted the normalized CO2 emissions data to visualize the trend over time.

MODEL TRAINING AND EVALUATION

We trained an XGBoost regressor on the training set and evaluated its performance using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) metrics.

We also trained and compared several other regression models including Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM).

FEATURE ENGINEERING

We created additional time series features such as quarter, month, and year to capture seasonal and temporal patterns in the data.

DATA SPLITTING

The dataset was split into training and testing sets using an 80-20 split ratio to train our models and evaluate their performance.

VISUALIZATION

We created additional time series features such as quarter, month, and year to capture seasonal and temporal patterns in the data.

RESULTS

- The Decision Tree model had the best performance among the models we tried, with an R2 score of 0.14785, indicating some degree of correlation between predicted and actual values. It also had the lowest MSE (0.00391) and a relatively low MAE (0.05151).

- Other models, such as **Random Forest**, **Lasso Regression**, **Support Vector Machine**, **Linear Regression**, and **Ridge Regression**, showed negative R2 scores, indicating poor performance and suggesting that these models did not fit the data well.

- The Random Forest model had an R2 score of -0.49203 with an MSE of 0.00684 and an MAE of 0.07362, performing better than other poorly performing models but still not effectively capturing the trend.

- **Lasso Regression**, **Support Vector Machine**, **Linear Regression**, and **Ridge Regression** all performed poorly, with significantly negative R2 scores and higher MSE and MAE values.

	REGRESSION NAME	R2_SCORE	MSE	MAE
0	Decision Tree	0,14785	0,00391	0,05151
1	Random Forest	-0,49203	0,00684	0,07362
2	Lasso Regression	-5,08395	0,02790	0,15820
3	Support Vector Machine	-8,18545	0,04212	0,19447
4	Linear Regression	-11,30931	0,05644	0,22738
5	Ridge Regression	-11,32269	0,05650	0,22756

DATASET REFERENCE

The dataset used in this project was obtained from [**OECD**], which contains historical data on CO2 emissions from cars worldwide. The data includes time series information and CO2 emission values, essential for building and evaluating predictive models.