1. Sprint 3

In this sprint, we want to train the baseline models with the preprocessed data and use it to predict the CO2 emissions from different types of cars. The target feature to estimate is CO2 (g/km). In the following sections, we present the results of baseline models. We will present the results of two scenarios: with and without using dimensionality reduction techniques.

1.1. Linear Regression

As the first model, we trained the Linear Regression model.

1.1.1. Linear Regression without Dimensionality Reduction

The results of applying Linear Regression without reducing the dimension are:

- Mean Squared Error: 2.6485711266259274e-05
- Mean Absolute Error (MAE): 0.0025955619001083283
- Root Mean Squared Error (RMSE): 0.005146427038855139
- R-squared (R2): 0.9944408953741078

A low MSE, MAE, and RMSE indicate that the model's predictions are close to the actual values, while a high R-squared value suggests that the model explains a significant proportion of the variance in the target variable. Overall, these results indicate that the model performs well in predicting CO2 emissions.

1.1.2. Linear Regression with Dimensionality Reduction

We applied PCA to reduce the dimensionality of the dataset. The results of evaluation metrics are:

- Mean Squared Error (MSE): 0.0005988825887608515
- Mean Absolute Error (MAE): 0.017800248244260315
- Root Mean Squared Error (RMSE): 0.024472077736899488
- R-squared (R2): 0.8743001108757107

Reducing the dimensionality with PCA while retaining 95% of the variance has led to faster model application, which is a common advantage of dimensionality reduction techniques. However, the evaluation metrics show that the model's performance slightly decreased after applying PCA compared to the previous model without dimensionality reduction. The mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE) have increased, indicating higher prediction errors. Additionally, the R-squared (R2) value decreased, suggesting that the model explains less variance in the target variable.

1.2. Ridge Regression

Ridge Regression is a type of linear regression that incorporates regularization to address high correlation between predictor variables and reduce the model's sensitivity to noisy input data. The results of applying this model are presented in the next sections.

1.2.1. Ridge Regression without Dimensionality Reduction

The results of applying Ridge Regression without reducing the dimension are:

- Mean Squared Error (MSE): 2.9221645265128422e-05
- Mean Absolute Error (MAE): 0.0028743076853023158
- Root Mean Squared Error (RMSE): 0.005405704881431137
- R-squared (R2): 0.9938666482566205

This model seems to perform well based on these metrics, with low errors and a high R-squared value indicating a good fit to the data.

To identify the optimal value of alpha, we used the cross validation. These are the results of the application of the model with optimal value for alpha:

Optimal Alpha: 0.1

- Mean Squared Error (MSE): 2.7492705385780724e-05
- Mean Absolute Error (MAE): 0.0026688016450271846
- Root Mean Squared Error (RMSE): 0.005243348680545737
- R-squared (R2): 0.9942295366678302

It seems that using the optimal alpha value has resulted in improvements in most of the evaluation metrics. The Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) values are slightly better with the optimal alpha, indicating that the model's performance improved after tuning the regularization parameter.

1.2.2. Ridge Regression with Dimensionality Reduction

We used PCA to reduce the dimensionality of the dataset. The results of the application of Ridge Regression model after reducing the dimensions are:

Optimal Alpha: 1

- Mean Squared Error (MSE): 0.0005945705163835583
- Mean Absolute Error (MAE): 0.017751162611685575
- Root Mean Squared Error (RMSE): 0.024383816690246796
- R-squared (R2): 0.8752051747895626

It appears that before applying PCA, the Ridge Regression model with an optimal alpha value of 0.1 achieved slightly better performance compared to after applying PCA. The R-squared value was higher (approximately 0.99), indicating that the model explained more variance in the target variable before PCA. Additionally, the MSE, MAE, and RMSE values were lower, indicating lower prediction errors.

This suggests that the original features without dimensionality reduction might have contained more predictive power for estimating CO2 emissions. However, it's essential to consider the trade-offs between model performance and computational efficiency, as dimensionality reduction techniques like PCA can help reduce the complexity of the model and speed up computation, especially with larger datasets.

1.3. Lasso Regression

Lasso Regression, also known as L1 regularization, is a linear regression technique that adds a penalty term to the ordinary least squares objective function. This penalty term is the absolute value of the coefficients multiplied by a regularization parameter (alpha), which controls the strength of regularization.

The main objective of Lasso Regression is to minimize the sum of the squared residuals between the observed and predicted values, similar to ordinary linear regression. However, it also aims to minimize the sum of the absolute values of the coefficients, thereby encouraging sparsity in the coefficient matrix. This means that Lasso Regression tends to produce models with fewer coefficients, effectively performing feature selection by shrinking less important coefficients towards zero.

The regularization parameter, alpha, controls the balance between fitting the data well and keeping the model simple. Higher values of alpha result in more regularization, leading to more coefficients being set to zero and a simpler model.

In summary, Lasso Regression is a useful technique for feature selection and regularization in linear regression models, particularly when dealing with high-dimensional datasets with potentially redundant or irrelevant features.

Using Lasso Regression alongside Ridge Regression is a common practice, especially when dealing with high-dimensional datasets or when you want to perform feature selection. Lasso Regression tends to produce sparse models by setting some coefficients to zero, effectively performing feature selection, whereas Ridge Regression tends to shrink the coefficients towards zero without necessarily setting them exactly to zero.

The results of the application of this model on our dataset are:

- Mean Squared Error (MSE): 0.004764564350626007
- Mean Absolute Error (MAE): 0.05109835557138939
- Root Mean Squared Error (RMSE): 0.06902582379534494
- R-squared (R2): -3.777341135546841e-05

These results show a high mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE), as well as a negative R-squared value. This indicates that the model is performing poorly and may not be capturing the relationship between the features and the target variable effectively.

We tried to tune the value of alpha. The results of this model with the optimal value of alpha are:

Optimal Alpha: 0.01

- Mean Squared Error (MSE): 0.0029877575268617627
- Mean Absolute Error (MAE): 0.0374890930641223
- Root Mean Squared Error (RMSE): 0.05466038352281991
- R-squared (R2): 0.3728974645366506

These results show that the performance of the Lasso Regression model with the optimal alpha value (0.01) improved compared to the previous application. However, the model's performance still seems suboptimal, as indicated by the relatively low R-squared value (0.3729) and the error metrics.

Therefore, this model is not a good model for our project.

1.4. ElasticNet Regression

ElasticNet Regression is a combination of Ridge and Lasso regression techniques, which combines their penalties. The results of applying this method to the dataset are:

- Mean Squared Error (MSE): 0.004764564350626007
- Mean Absolute Error (MAE): 0.05109835557138939
- Root Mean Squared Error (RMSE): 0.06902582379534494
- R-squared (R2): -3.777341135546841e-05

The results show that the values of MSE, MAE, and RMSE are low. But, the negative value of R-squared indicates that the model performs worse than a horizontal line. This suggests that the model is not able to explain the variance in the data and is performing poorly.

We tried to perform hyperparameter tuning for ElasticNet regression using cross-validation with GridSearchCV. The results are:

- Best Hyperparameters: {'alpha': 0.1, 'l1_ratio': 0.1}
- Mean Squared Error (MSE): 0.0033337131647687627
- R-squared (R2): 0.30028459159136967

With these hyperparameters, the ElasticNet model achieved an MSE of approximately 0.0033 and an R2 value of approximately 0.3003 on the test set. Overall, while the model's performance improved compared to the previous results, it still may not be satisfactory, especially considering the relatively low R-squared value.

Therefore, this model is not a good model for our project.

1.5. Decision Trees

Decision Trees are a popular algorithm for regression tasks; therefore, we decided to apply this model on our dataset. We applied this model with and without dimensionality reduction. The results are presented in the next two sections.

1.5.1. Decision Trees without Dimensionality Reduction

The results of this model are:

- Mean Squared Error (MSE): 5.0408947205889875e-06
- R-squared (R2): 0.9989419630502595
- Mean Absolute Error (MAE): 0.00016894822536393775
- Root Mean Squared Error (RMSE): 0.0022451936933344943
- Explained Variance Score: 0.9989419815252438

A very low MSE, high R-squared, and small MAE and RMSE indicate that this model is performing exceptionally well and is able to make accurate predictions with very small errors. Additionally, the explained variance score of nearly 1 suggests that the model is able to explain almost all of the variance in the target variable. Overall, these metrics indicate **that the Decision Tree Regression model is highly effective for the task at hand.**

1.5.2. Decision Trees with Dimensionality Reduction

The results of this model are:

- Mean Squared Error (MSE): 7.293080520500923e-05
- R-squared (R2): 0.9846925018358221
- Mean Absolute Error (MAE): 0.0012739928590183088
- Root Mean Squared Error (RMSE): 0.008539953466208656
- Explained Variance Score: 0.9846934559572341

These results show that the Decision Tree model without dimensionality reduction performed better across all metrics compared to the one with dimensionality reduction.

It should be noted that applying PCA before a decision tree (or ensemble methods like Random Forest) is less common compared to using it with linear models. Decision trees, including Random Forest, are capable of handling high-dimensional data and nonlinear relationships without the need for dimensionality reduction techniques like PCA.

1.6. Random Forest

Random Forest is a popular ensemble learning technique that combines multiple decision trees to improve predictive performance and reduce overfitting. It's known for its robustness and ability to handle large datasets with high dimensionality.

The results of applying this model are:

- Mean Squared Error (MSE): 4.391058709938986e-06
- R-squared (R2): 0.9990783575890566
- Mean Absolute Error (MAE): 0.00020664446344440724
- Root Mean Squared Error (RMSE): 0.0020954853160876567
- Explained Variance Score: 0.9990784006100025

These results show that **<u>the Random Forest model seems to perform very well</u>**, achieving low values for MSE, MAE, and RMSE, and a high value for R2, indicating a good fit to the data. The Explained Variance Score is also very high, suggesting that the model explains most of the variance in the target variable.

	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	R-squared (R2)
Linear Regression	2.5862319913672156e-05	0.002660471599808402	0.005085500950120072	0.9945900491367343
Linear Regression with PCA	0.0005918952215822852	0.017638139928485598	0.024328896842690694	0.8761857375652871
Ridge Regression	2.7046578167122842e-05	0.002802543003414635	0.005200632477605281	0.9943423227540289
Ridge Regression with PCA	0.0005877155529261629	0.017591253813664616	0.024242845396655957	0.8770600520942942
Lasso Regression	0.0029816903391121316	0.0374522644403457	0.054604856369302276	0.376281854144763
ElasticNet Regression	0.0033313645142531365	0.051255177801859994	0.06914252535022192	0.3031360531500995
Decision Trees	4.8277542758872035e-06	0.00017087075484179201	0.002197215118254743	0.9989901171472764
Decision Trees with PCA	7.538883977648022e-05	0.001315519499204571	0.008682674690236886	0.9842299561605161
Random Forest	4.253333370564166e-06	0.00020071959446116072	0.0020623611154606666	0.9991102760844097