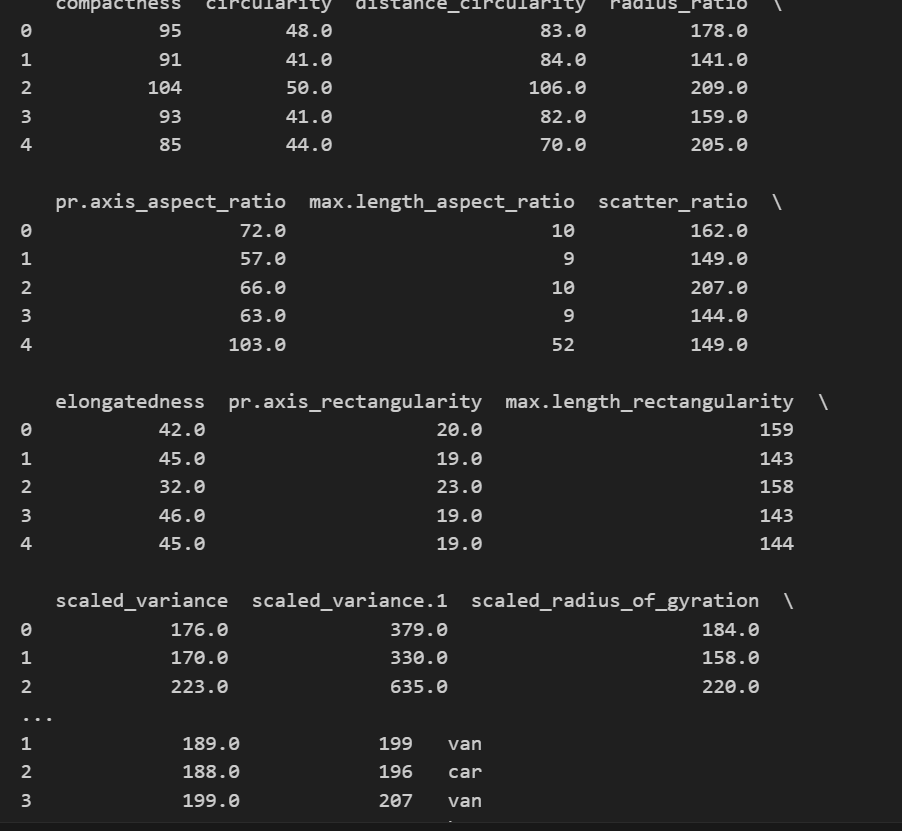
**Documentation**

1. **Dataset Information**

**Link to dataset:** <https://www.kaggle.com/datasets/pritech/vehicle-silhouettes?resource=download>

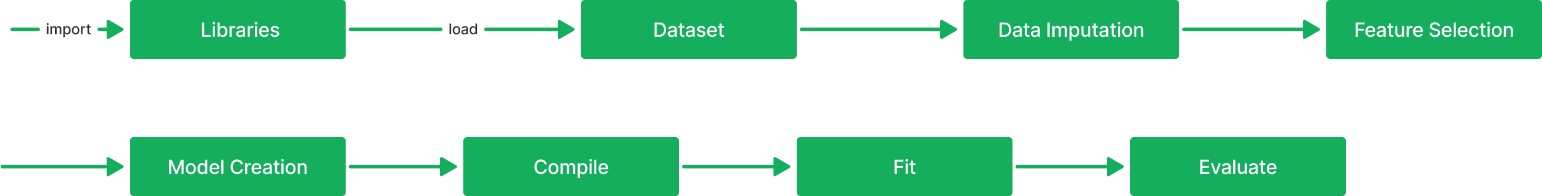
This dataset contains 18 features which are the properties of the vehicle’s silhouettes. The target variable is the class which it belongs either car, bus, or van.

1. **Dataset with Imputed values**

****

There are no more missing values in the dataset and they are now imputed.

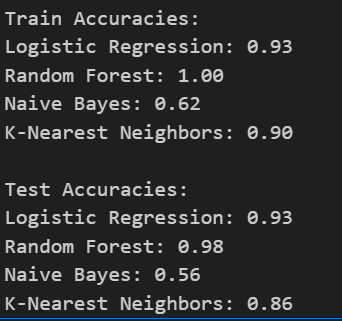
1. **Flow Chart of Processes**

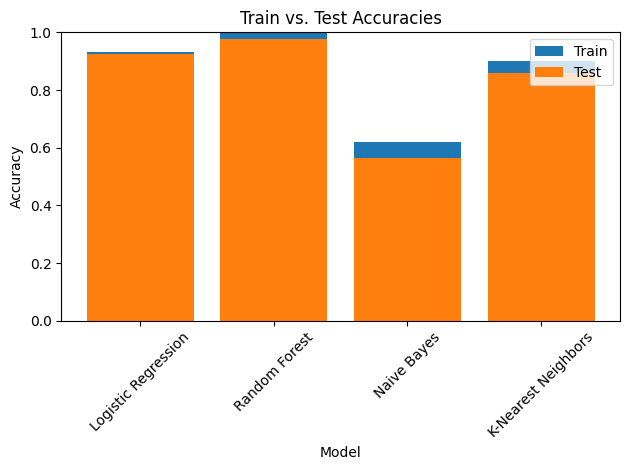


The process is the same for test one, except here we have to perform features selection. Which means we have to reduce some features in the dataset and see if it improves our model.

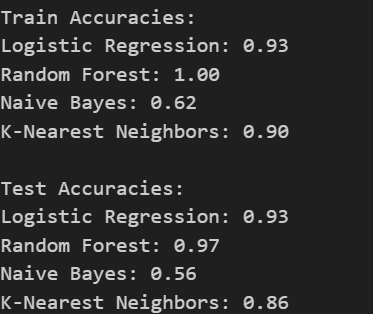
1. Results

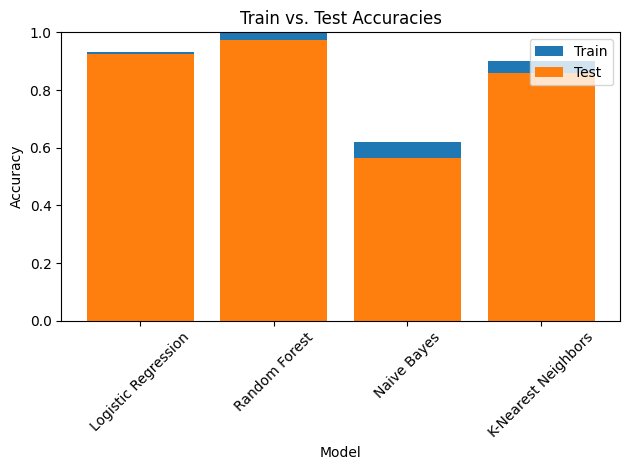
Before feature selection:





After feature selection:





The algorithm which showed promising results is the Random Forest Classifier with 97% percent accuracy in the testing set. It also shows a little sign of overfitting since there is little difference with accuracy in the training and testing sets. For this analysis, the dataset which we included all features showed better results with 98% for Random Forest Classifier. However, there are many other feature reduction techniques that can be done which might significantly improve the models performance in addition with hyperparameter tuning and using other machine learning algorithms. te