Data set details

1. Dataset file: 'cars_class.csv'
2. This is a multi-class classification data set.
3. The data set has 719 samples.
4. There are 18 numerical features.
5. The target variable is the class of the car which may be one of: 0 –bus, 1 – Opel Manta, 2 – Saab, 3 – Van.

Description of attributes

Comp: Compactness
Circ: Circularity
D.Circ: Distance Circularity
Rad.Ra: Radius ratio
Pr.Axis.Ra: pr.axis aspect ratio
Max.L.Ra: max.length aspect ratio
Scat.Ra: scatter ratio
Elong: elongatedness
Pr.Axis.Rect: pr.axis rectangularity
Max.L.Rect: max.length rectangularity
Sc.Var.Maxis: scaled variance along major axis
Sc.Var.maxis: scaled variance along minor axis
Ra.Gyr: scaled radius of gyration
Skew.Maxis: skewness about major axis
Skew.maxis: skewness about minor axis
Kurt.maxis: kurtosis about minor axis
Kurt.Maxis: kurtosis about major axis
Holl.Ra: hollows ratio

Tasks

The following are the broad tasks to be performed:

- Load the data into your code
- Keep 20% of the data aside as test data.
- Apply the required preprocessing techniques on the data.
- Apply different machine learning techniques.
- Tune the parameters of models as required.
- Build a model named 'final_model'
- Report the following metrics using 'final_model' on the test data:
    - Accuracy
    - F1-score
    - Display the confusion matrix
- Report the importance of different features.
- **Prepare a brief report documenting the key steps involved along with the reasoning behind different decisions that you made in these steps.**

Notes

- Write textual comments in your code wherever appropriate.
- **Note:** As a data scientist, you might need to make several decisions while exploring the data. feel free to do that and mention the same in the report.
- Be as innovative and creative as possible. Visualize data wherever required.