Methodology

* The 3 classifiers used
* Ensemble pipeline
* Other models considered
* Hyper-parameter tuning

*[7. Provide a high-level description of Machine Learning models – association rules and random forest to predict.*

*8: Describe the entire data set and identify your observations from this.]*

Overview

* Objective
* Methodology

Dataset

* How many features
* Size of the dataset
* Multiple files
* What kind of data – numerical or character
* Balanced or imbalanced – what is the distribution
* Distribution of Training set, validation set, testing set
* Missing data and Preprocessing challenges

*[2. Perform exploratory analysis on the data and describe your understanding of the data.*

*3. Perform data wrangling / pre-processing. E.g., missing data, normalization, discretization, etc.,]*

Results

* Table for the evaluation metric for each ML technique used
* Plot of the curves
* Conclusion

*[9. Plot the histogram of various data sets (at least for 3 parameters i.e. year in which the car was bought., selling price, showroom price)*

*10. Present the conclusions/results (Answers for the given objectives) in the format shared.]*

Feature Engineering Techniques

* Features removed
* Feature creation
* Feature ranking
* Class imbalance treatment
* Any other

*[4. Apply any two feature selection engineering techniques.*

*5. Compare the two selected feature engineering techniques.*

*6. Plot top 5, 6, and 8 features.]*