Methodology

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

Overview

* To predict health insurance owners who will be interested in vehicle insurance
* Methodology

Dataset

* How many features
* Size of the dataset  
    
  Healthcare Analytics: Train Set – 75278; Test Set – 35249

Vehicle Insurance: Train Set – 381109; Test Set – 127037

* 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

Results

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

Feature Engineering Techniques

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