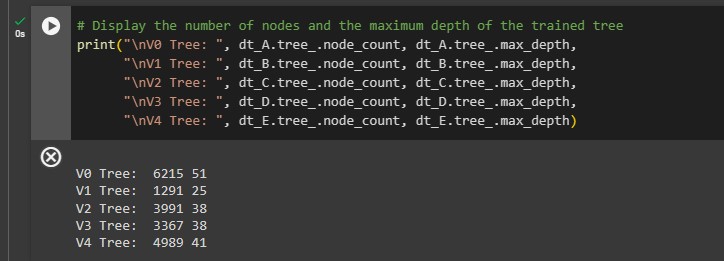
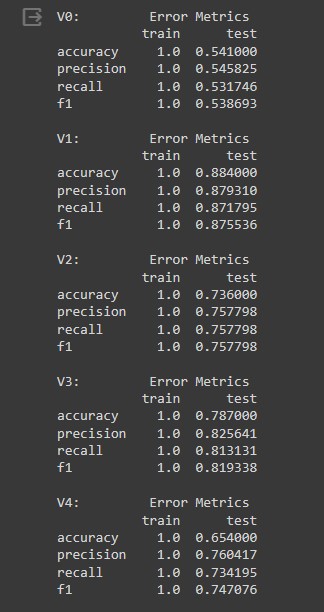
# Machine Learning: Driver Overtake Scenario

## Data Pre-Processing:

* The Scenario focuses on making and training a machine learning system capable of accurately predicting whether a driver should overtake the car in front or not, given a limited set of situational data, courtesy of the black-box routine.
* Data collection involves generating a .csv file containing several thousand records, in this instance: 10,000 per file.
* Pre-processing:
  + Involves converting the “Success” column from text values TRUE/FALSE to integer values 1/0.
  + Extracting the feature columns into a data subset
  + Performing data-splitting operations such as a stratified-shuffle-split to generate train and test data subsets
* Checking the data-processing is working as expected

## Data Processing/Model Testing:



Decision Tree used for similarity to real-driver thought processes, minus clear numerical backing.

A Decision Tree Classifier is a useful algorithm that generate a flowchart-like tree structure where internal nodes refer to tests on attributes and branches represent outcomes.

A Decision Tree Classifier is constructed by recursively splitting training data into subsets until a stop criteria is met.

Grid Search Cross-Validation is a technique for tuning hyper-parameters (things such as max\_depth, min\_sample\_split\_size, etc)

The Data used in this scenario is generated by a black-box routine, governed by intentionally unknown algorithms and variables. Its usage has no real ethical or social context due to the black-box routine’s generality and non-specificity. Economic and legal considerations, are middling. Without further tuning, or comparison to data gathered/generated in the real-world, the black-box routine serves as a good training and testing tool. It should not inform legally significant decisions such as whether a car needs seatbelts or something.

## Conclusion:

Decision Tree Classifier, supported with Grid Search Cross-Validation, offers an effective method of dealing with the scenario.

## Github Link:

https://github.com/Jacob-L-ARU-P/ML