

## Ζήτημα 1ο

- Επιλογή χαρακτηριστικών (οπτικοί έλεγχοι , select attributes, Remove useless)
- Από το select attributes οι πιο σημαντικές μεταβλητές είναι οι (1, 2, 3, 6, 20)
- Με τους οπτικούς ελέγχους παρατηρώ ότι όλες οι μεταβλητές είναι σημαντικές
- Η Remove useless για ποσοστό 98% δεν διαγράφει καμία μεταβλητή
- Αποφασίζω να κρατήσω όλες τις μεταβλητές
- Δοκιμάζω κάποιους απλούς αλγόριθμους , τα καλύτερα αποτελέσματα δίνουν οι PART, BFTree, j48, j48graft, RandomForest , LADTree, NBTree
- Δοκιμάζω κάποιους μετά-αλγόριθμους με βάση τους πιο σημαντικούς απλούς αλγόριθμους
- Επιλέγω τον αλγόριθμο: Bagging → PART

## Ζήτημα 2ο

Τρέχουμε κάποιους αλγόριθμους

Ο OneR δίνει τους παρακάτω κανόνες

SCALED VARIANCE\_MINOR:

< 289.5	-> van
< 298.5	-> saab
< 307.5	-> van
< 312.5	-> bus
< 318.5	-> van
< 347.5	-> bus
< 380.0	-> van
< 390.5	-> saab
< 418.5	-> opel
< 431.5	-> bus
< 465.5	-> saab
< 475.0	-> opel
< 488.0	-> bus
< 528.5	-> saab
< 575.5	-> opel
< 588.0	-> saab
< 603.0	-> opel
< 621.5	-> saab
< 634.0	-> opel
< 640.5	-> saab
< 657.5	-> opel
< 677.5	-> saab
< 684.5	-> opel
< 705.5	-> saab
< 708.5	-> opel
< 721.5	-> saab
< 761.5	-> opel
>= 761.5	-> bus

(518/846 instances correct)

από όπου ανάλογα με την τιμή του SCALED VARIANCE\_MINOR προκύπτει ο τύπος του οχήματος.

Ο PART δίνει κάποιους κανόνες από τους οποίους οι πιο σημαντικοί είναι

ELONGATEDNESS <= 41 AND  
MAX.LENGTH ASPECT RATIO <= 7 AND  
COMPACTNESS > 95: bus (70.0/1.0)

ELONGATEDNESS <= 41 AND  
MAX.LENGTH ASPECT RATIO > 7 AND  
SCALED VARIANCE\_MINOR <= 721 AND  
HOLLOWS RATIO <= 195 AND  
MAX.LENGTH ASPECT RATIO > 8 AND  
KURTOSIS ABOUT\_MINOR > 185: opel (23.0)

ELONGATEDNESS <= 41 AND  
MAX.LENGTH ASPECT RATIO > 7 AND  
SCALED VARIANCE\_MINOR <= 721 AND  
COMPACTNESS <= 109 AND  
SKEWNESS ABOUT\_MAJOR > 67 AND  
COMPACTNESS > 106: saab (24.0/1.0)

ELONGATEDNESS <= 41 AND  
MAX.LENGTH ASPECT RATIO > 7 AND  
SCALED VARIANCE\_MINOR > 721: opel (23.0/1.0)

ELONGATEDNESS <= 41 AND  
PR.AXIS ASPECT RATIO <= 68 AND  
COMPACTNESS > 109: saab (16.0)

ELONGATEDNESS <= 41 AND  
PR.AXIS ASPECT RATIO <= 68 AND  
MAX.LENGTH ASPECT RATIO > 7 AND  
PR.AXIS ASPECT RATIO <= 57 AND  
HOLLOWS RATIO > 197: opel (9.0)

ELONGATEDNESS <= 41 AND  
PR.AXIS ASPECT RATIO <= 68 AND  
MAX.LENGTH ASPECT RATIO > 7 AND  
SKEWNESS ABOUT\_MINOR > 12 AND  
KURTOSIS ABOUT\_MAJOR > 3 AND  
KURTOSIS ABOUT\_MINOR <= 198: saab (16.0)

ELONGATEDNESS <= 41 AND  
PR.AXIS ASPECT RATIO <= 68 AND  
SKEWNESS ABOUT\_MINOR > 10 AND  
SKEWNESS ABOUT\_MAJOR <= 69: opel (16.0/1.0)

MAX.LENGTH ASPECT RATIO > 8 AND  
HOLLOWS RATIO > 189 AND  
SKEWNESS ABOUT\_MAJOR > 63: van (107.0/2.0)

SCALED VARIANCE\_MINOR > 309 AND  
DISTANCE CIRCULARITY <= 76 AND  
SKEWNESS ABOUT\_MINOR <= 10 AND  
SKEWNESS ABOUT\_MAJOR > 64 AND  
CIRCULARITY > 41: bus (107.0/1.0)

PR.AXIS ASPECT RATIO > 63 AND  
ELONGATEDNESS <= 47 AND  
MAX.LENGTH ASPECT RATIO <= 7 AND  
KURTOSIS ABOUT\_MAJOR <= 23 AND  
SCALED RADIUS OF GYRATION <= 185: bus (35.0)

SCALED VARIANCE\_MAJOR <= 173 AND  
COMPACTNESS <= 82 AND  
COMPACTNESS <= 81 AND  
KURTOSIS ABOUT\_MAJOR > 4: opel (12.0)

SCALED VARIANCE\_MAJOR <= 173 AND  
COMPACTNESS > 82 AND  
KURTOSIS ABOUT\_MAJOR <= 19 AND  
SCATTER RATIO <= 142 AND  
PR.AXIS RECTANGULARITY <= 17 AND  
SKEWNESS ABOUT\_MINOR <= 18: van (30.0/1.0)

SCALED VARIANCE\_MAJOR > 173 AND  
PR.AXIS ASPECT RATIO <= 69 AND  
COMPACTNESS > 91: saab (8.0/1.0)

SCALED VARIANCE\_MAJOR <= 173 AND  
MAX.LENGTH RECTANGULARITY > 138 AND  
PR.AXIS RECTANGULARITY <= 18 AND  
ELONGATEDNESS > 47: van (33.0)

PR.AXIS ASPECT RATIO > 65 AND  
PR.AXIS ASPECT RATIO <= 70: van (4.0)

PR.AXIS ASPECT RATIO <= 68 AND  
MAX.LENGTH ASPECT RATIO > 7 AND  
KURTOSIS ABOUT\_MAJOR <= 17 AND  
PR.AXIS ASPECT RATIO > 58: van (12.0)

ELONGATEDNESS > 46 AND  
MAX.LENGTH ASPECT RATIO > 4 AND  
SCALED RADIUS OF GYRATION <= 124 AND  
KURTOSIS ABOUT\_MAJOR <= 28 AND  
MAX.LENGTH RECTANGULARITY > 124: van (9.0)

PR.AXIS ASPECT RATIO <= 68 AND  
MAX.LENGTH ASPECT RATIO <= 4: saab (7.0)

PR.AXIS ASPECT RATIO <= 68 AND  
RADIUS RATIO <= 120 AND  
CIRCULARITY <= 38: saab (6.0/1.0)

Απο τους παρακάτω κανόνες μπορούμε να βγάλουμε κάποια συμπεράσματα όπως π.χ. ο κανόνας:

ELONGATEDNESS <= 41 AND  
MAX.LENGTH ASPECT RATIO <= 7 AND  
COMPACTNESS > 95: bus (70.0/1.0)

μας λέει ότι αν έχουμε

ELONGATEDNESS <= 41  
και  
MAX.LENGTH ASPECT RATIO <= 7  
και  
COMPACTNESS > 95  
τότε ο τύπος του οχήματος είναι bus.

Οι τιμές στις παρενθέσεις δείχνουν σε πόσες περιπτώσεις ο κανόνας δίνει σωστές προβλέψεις και σε πόσες λάθος (Σ/Λ). Παρατηρούμε ότι ο προηγούμενος κανόνας δίνει 70 σωστές προβλέψεις και μόνο 1 λάθος.

O Decision Stump δίνει:

Classifications

ELONGATEDNESS <= 41.5 : saab  
ELONGATEDNESS > 41.5 : van  
ELONGATEDNESS is missing : bus

Class distributions

ELONGATEDNESS <= 41.5  
opel saab bus van  
0.38481675392670156 0.387434554973822 0.22774869109947643 0.0  
ELONGATEDNESS > 41.5  
opel saab bus van  
0.14008620689655170.14870689655172414 0.28232758620689660.42887931034482757  
ELONGATEDNESS is missing  
opel saab bus van  
0.25059101654846333 0.2565011820330969 0.25768321513002360.23522458628841608