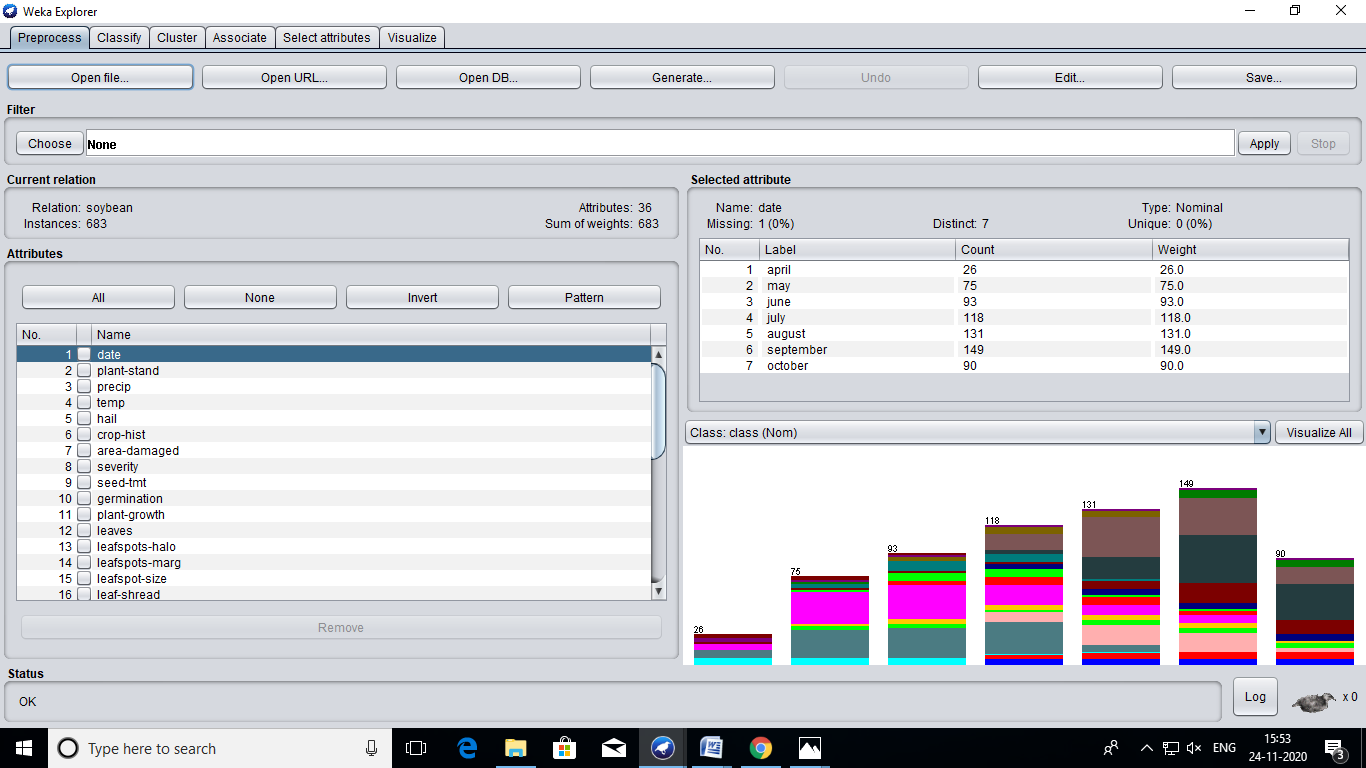
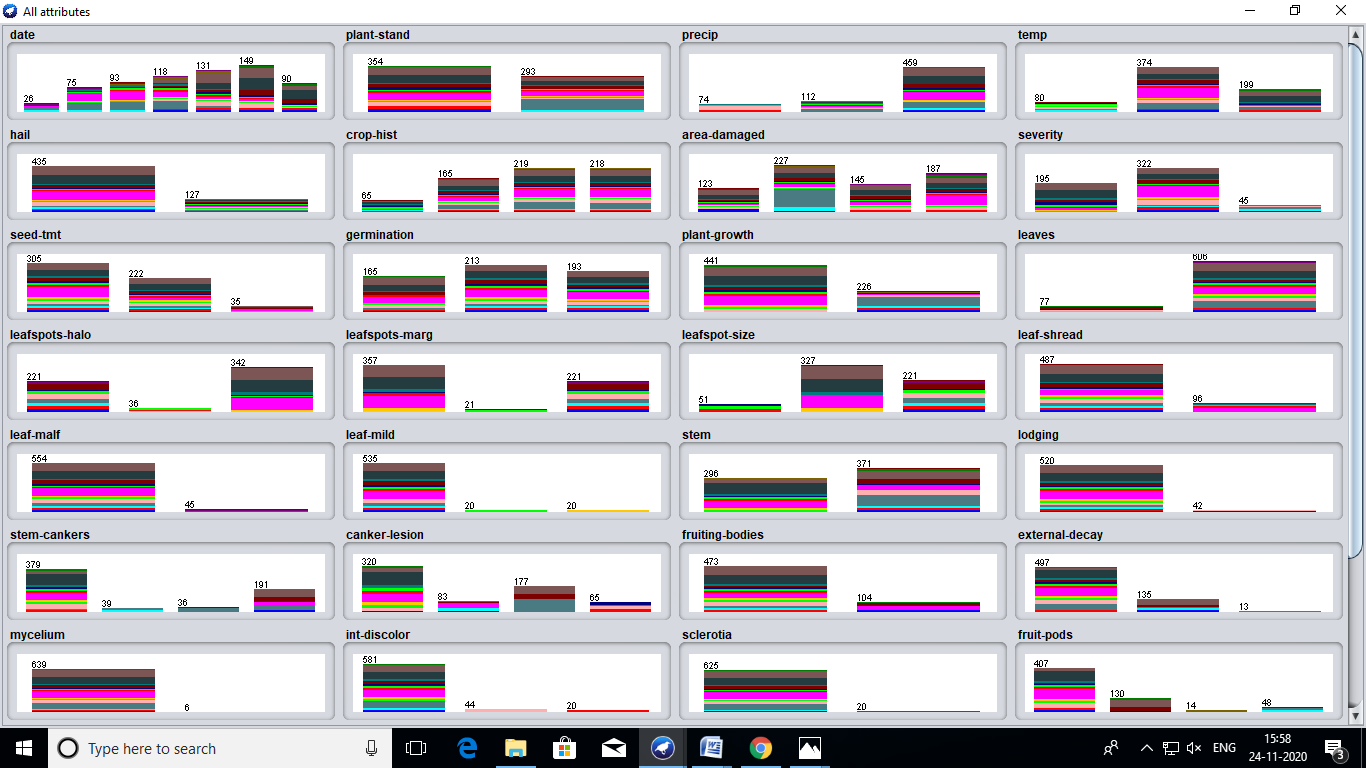
1. **Write a program to implement dbscan algorithm**





=== Run information ===

Scheme: weka.clusterers.MakeDensityBasedClusterer -M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: soybean

Instances: 683

Attributes: 36

date

plant-stand

precip

temp

hail

crop-hist

area-damaged

severity

seed-tmt

germination

plant-growth

leaves

leafspots-halo

leafspots-marg

leafspot-size

leaf-shread

leaf-malf

leaf-mild

stem

lodging

stem-cankers

canker-lesion

fruiting-bodies

external-decay

mycelium

int-discolor

sclerotia

fruit-pods

fruit-spots

seed

mold-growth

seed-discolor

seed-size

shriveling

roots

class

Test mode: evaluate on training data

=== Clustering model (full training set) ===

MakeDensityBasedClusterer:

Wrapped clusterer:

kMeans

======

Number of iterations: 10

Within cluster sum of squared errors: 6209.0

Initial starting points (random):

Cluster 0: september,normal,gt-norm,norm,yes,same-lst-two-yrs,whole-field,severe,fungicide,80-89,norm,abnorm,no-yellow-halos,w-s-marg,gt-1/8,absent,absent,absent,abnorm,yes,absent,tan,present,firm-and-dry,absent,none,absent,norm,absent,norm,absent,absent,norm,absent,norm,brown-spot

Cluster 1: june,lt-normal,gt-norm,norm,yes,same-lst-sev-yrs,whole-field,pot-severe,none,lt-80,norm,abnorm,no-yellow-halos,w-s-marg,gt-1/8,present,absent,absent,norm,yes,absent,dna,absent,absent,absent,none,absent,norm,absent,norm,absent,absent,norm,absent,norm,brown-spot

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1

(683.0) (261.0) (422.0)

==========================================================================================

date september july september

plant-stand normal lt-normal normal

precip gt-norm gt-norm gt-norm

temp norm norm norm

hail yes yes yes

crop-hist same-lst-two-yrs same-lst-two-yrs same-lst-sev-yrs

area-damaged low-areas low-areas whole-field

severity pot-severe pot-severe pot-severe

seed-tmt none none none

germination 80-89 80-89 80-89

plant-growth norm abnorm norm

leaves abnorm abnorm abnorm

leafspots-halo no-yellow-halos absent no-yellow-halos

leafspots-marg w-s-marg dna w-s-marg

leafspot-size gt-1/8 dna gt-1/8

leaf-shread absent absent absent

leaf-malf absent absent absent

leaf-mild absent absent absent

stem abnorm abnorm norm

lodging yes yes yes

stem-cankers absent above-sec-nde absent

canker-lesion dna dk-brown-blk dna

fruiting-bodies absent absent absent

external-decay absent absent absent

mycelium absent absent absent

int-discolor none none none

sclerotia absent absent absent

fruit-pods norm norm norm

fruit-spots absent absent absent

seed norm norm norm

mold-growth absent absent absent

seed-discolor absent absent absent

seed-size norm norm norm

shriveling absent absent absent

roots norm norm norm

class brown-spot phytophthora-rot alternarialeaf-spot

Fitted estimators (with ML estimates of variance):

Cluster: 0 Prior probability: 0.3825

Attribute: date

Discrete Estimator. Counts = 21 36 39 57 43 38 34 (Total = 268)

Attribute: plant-stand

Discrete Estimator. Counts = 103 160 (Total = 263)

Attribute: precip

Discrete Estimator. Counts = 37 46 181 (Total = 264)

Attribute: temp

Discrete Estimator. Counts = 52 151 61 (Total = 264)

Attribute: hail

Discrete Estimator. Counts = 221 42 (Total = 263)

Attribute: crop-hist

Discrete Estimator. Counts = 28 59 103 75 (Total = 265)

Attribute: area-damaged

Discrete Estimator. Counts = 41 145 46 33 (Total = 265)

Attribute: severity

Discrete Estimator. Counts = 28 198 38 (Total = 264)

Attribute: seed-tmt

Discrete Estimator. Counts = 179 76 9 (Total = 264)

Attribute: germination

Discrete Estimator. Counts = 47 156 61 (Total = 264)

Attribute: plant-growth

Discrete Estimator. Counts = 85 178 (Total = 263)

Attribute: leaves

Discrete Estimator. Counts = 59 204 (Total = 263)

Attribute: leafspots-halo

Discrete Estimator. Counts = 192 1 71 (Total = 264)

Attribute: leafspots-marg

Discrete Estimator. Counts = 71 1 192 (Total = 264)

Attribute: leafspot-size

Discrete Estimator. Counts = 1 72 191 (Total = 264)

Attribute: leaf-shread

Discrete Estimator. Counts = 261 2 (Total = 263)

Attribute: leaf-malf

Discrete Estimator. Counts = 242 21 (Total = 263)

Attribute: leaf-mild

Discrete Estimator. Counts = 253 10 1 (Total = 264)

Attribute: stem

Discrete Estimator. Counts = 12 251 (Total = 263)

Attribute: lodging

Discrete Estimator. Counts = 233 30 (Total = 263)

Attribute: stem-cankers

Discrete Estimator. Counts = 83 40 36 106 (Total = 265)

Attribute: canker-lesion

Discrete Estimator. Counts = 49 46 131 39 (Total = 265)

Attribute: fruiting-bodies

Discrete Estimator. Counts = 208 55 (Total = 263)

Attribute: external-decay

Discrete Estimator. Counts = 175 75 14 (Total = 264)

Attribute: mycelium

Discrete Estimator. Counts = 256 7 (Total = 263)

Attribute: int-discolor

Discrete Estimator. Counts = 220 23 21 (Total = 264)

Attribute: sclerotia

Discrete Estimator. Counts = 242 21 (Total = 263)

Attribute: fruit-pods

Discrete Estimator. Counts = 169 50 1 45 (Total = 265)

Attribute: fruit-spots

Discrete Estimator. Counts = 119 12 40 1 94 (Total = 266)

Attribute: seed

Discrete Estimator. Counts = 230 33 (Total = 263)

Attribute: mold-growth

Discrete Estimator. Counts = 240 23 (Total = 263)

Attribute: seed-discolor

Discrete Estimator. Counts = 249 14 (Total = 263)

Attribute: seed-size

Discrete Estimator. Counts = 240 23 (Total = 263)

Attribute: shriveling

Discrete Estimator. Counts = 240 23 (Total = 263)

Attribute: roots

Discrete Estimator. Counts = 189 74 1 (Total = 264)

Attribute: class

Discrete Estimator. Counts = 21 21 21 89 23 10 1 6 1 1 6 45 1 1 9 1 1 17 5 (Total = 280)

Cluster: 1 Prior probability: 0.6175

Attribute: date

Discrete Estimator. Counts = 7 41 56 63 90 114 58 (Total = 429)

Attribute: plant-stand

Discrete Estimator. Counts = 289 135 (Total = 424)

Attribute: precip

Discrete Estimator. Counts = 39 68 318 (Total = 425)

Attribute: temp

Discrete Estimator. Counts = 30 255 140 (Total = 425)

Attribute: hail

Discrete Estimator. Counts = 337 87 (Total = 424)

Attribute: crop-hist

Discrete Estimator. Counts = 39 108 134 145 (Total = 426)

Attribute: area-damaged

Discrete Estimator. Counts = 84 85 101 156 (Total = 426)

Attribute: severity

Discrete Estimator. Counts = 169 247 9 (Total = 425)

Attribute: seed-tmt

Discrete Estimator. Counts = 249 148 28 (Total = 425)

Attribute: germination

Discrete Estimator. Counts = 120 171 134 (Total = 425)

Attribute: plant-growth

Discrete Estimator. Counts = 374 50 (Total = 424)

Attribute: leaves

Discrete Estimator. Counts = 20 404 (Total = 424)

Attribute: leafspots-halo

Discrete Estimator. Counts = 31 37 357 (Total = 425)

Attribute: leafspots-marg

Discrete Estimator. Counts = 372 22 31 (Total = 425)

Attribute: leafspot-size

Discrete Estimator. Counts = 52 341 32 (Total = 425)

Attribute: leaf-shread

Discrete Estimator. Counts = 328 96 (Total = 424)

Attribute: leaf-malf

Discrete Estimator. Counts = 398 26 (Total = 424)

Attribute: leaf-mild

Discrete Estimator. Counts = 392 12 21 (Total = 425)

Attribute: stem

Discrete Estimator. Counts = 286 138 (Total = 424)

Attribute: lodging

Discrete Estimator. Counts = 410 14 (Total = 424)

Attribute: stem-cankers

Discrete Estimator. Counts = 336 1 2 87 (Total = 426)

Attribute: canker-lesion

Discrete Estimator. Counts = 311 39 48 28 (Total = 426)

Attribute: fruiting-bodies

Discrete Estimator. Counts = 373 51 (Total = 424)

Attribute: external-decay

Discrete Estimator. Counts = 362 62 1 (Total = 425)

Attribute: mycelium

Discrete Estimator. Counts = 423 1 (Total = 424)

Attribute: int-discolor

Discrete Estimator. Counts = 401 23 1 (Total = 425)

Attribute: sclerotia

Discrete Estimator. Counts = 423 1 (Total = 424)

Attribute: fruit-pods

Discrete Estimator. Counts = 324 82 15 5 (Total = 426)

Attribute: fruit-spots

Discrete Estimator. Counts = 334 65 19 1 8 (Total = 427)

Attribute: seed

Discrete Estimator. Counts = 340 84 (Total = 424)

Attribute: mold-growth

Discrete Estimator. Counts = 378 46 (Total = 424)

Attribute: seed-discolor

Discrete Estimator. Counts = 372 52 (Total = 424)

Attribute: seed-size

Discrete Estimator. Counts = 386 38 (Total = 424)

Attribute: shriveling

Discrete Estimator. Counts = 407 17 (Total = 424)

Attribute: roots

Discrete Estimator. Counts = 395 14 16 (Total = 425)

Attribute: class

Discrete Estimator. Counts = 1 1 1 1 23 12 21 88 21 21 16 1 21 92 84 16 15 1 5 (Total = 441)

Time taken to build model (full training data) : 0.06 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 270 ( 40%)

1 413 ( 60%)

Log likelihood: -23.23893