=== Run information ===

Scheme: 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: yelp\_task2\_trainingset\_binaryclassification-weka.filters.unsupervised.attribute.NominalToString-Cfirst-weka.filters.unsupervised.attribute.StringToWordVector-R1-W2500-prune-rate-1.0-C-T-I-N1-L-S-stemmerweka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.NGramTokenizer -max 3 -min 1 -delimiters " \r\n\t.,;:\'\"()?!"-weka.filters.unsupervised.attribute.Reorder-R2-last,1-weka.filters.unsupervised.attribute.Reorder-R2-last,1-weka.filters.unsupervised.attribute.Reorder-R2-last,1-weka.filters.unsupervised.attribute.Reorder-R2-last,1-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelection.InfoGainAttributeEval-Sweka.attributeSelection.Ranker -T 0.0 -N -1

Instances: 27844

Attributes: 115

[list of attributes omitted]

Test mode: split 66% train, remainder test

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

kMeans

======

Number of iterations: 19

Within cluster sum of squared errors: 15831.133965860085

Initial staring points (random):

Cluster 0: {4 1.592602,7 2.041209,9 0.468058,10 1.809425,17 0.929551,33 0.448103,64 1.9024,80 2.174478,112 3}

Cluster 1: {17 1.364484}

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1

(27844) (4160) (23684)

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

airport 0.0376 0.0251 0.0398

bruschetta 0.0446 0.0365 0.0461

wine 0.0921 0.1303 0.0854

terminal 0.0146 0.0044 0.0164

vegas 0.0718 0.0803 0.0703

great 0.3495 0.3301 0.3529

mesa 0.0169 0.0091 0.0182

in vegas 0.0321 0.0331 0.032

postino 0.0224 0.0292 0.0212

food 0.3443 0.9853 0.2318

delicious 0.1465 0.2179 0.1339

rental 0.0182 0.0021 0.0211

love 0.2539 0.2425 0.2559

security 0.0194 0.0088 0.0213

henderson 0.0249 0.0297 0.024

the wine 0.0204 0.0475 0.0157

of wine 0.0281 0.0288 0.0279

place 0.3156 0.3927 0.302

insurance 0.0233 0 0.0274

rude 0.061 0.0368 0.0652

car 0.1106 0.0244 0.1258

customer 0.1229 0.0665 0.1328

manager 0.085 0.0859 0.0848

the airport 0.0149 0.0184 0.0143

the best 0.1862 0.1712 0.1888

office 0.0666 0.0155 0.0756

bus 0.0123 0.014 0.0119

greek 0.0227 0.0318 0.0211

called 0.0825 0.0505 0.0881

lgo 0.0158 0.0181 0.0155

customer service 0.0961 0.0439 0.1053

wine and 0.0237 0.0325 0.0221

bottle of 0.0195 0.0178 0.0198

good 0.3311 0.4369 0.3125

italian 0.0742 0.1191 0.0663

is great 0.0846 0.1446 0.074

excellent 0.1279 0.1695 0.1207

terrible 0.0536 0.0778 0.0494

bread 0.0933 0.1402 0.085

las vegas 0.0326 0.0365 0.0319

chicken 0.1859 0.2811 0.1691

call 0.0825 0.053 0.0877

sandwich 0.0938 0.0935 0.0938

wings 0.0942 0.0996 0.0933

grilled 0.038 0.052 0.0355

management 0.0341 0.0369 0.0336

told 0.1045 0.0677 0.1109

oil change 0.0237 0 0.0278

amazing 0.1489 0.1771 0.1439

company 0.0471 0.0217 0.0516

casino 0.0324 0.0308 0.0326

atmosphere 0.1092 0.1805 0.0967

credit 0.0342 0.0188 0.0369

cheese 0.128 0.1469 0.1247

las 0.0363 0.0424 0.0352

tomato 0.0356 0.045 0.0339

deli 0.0361 0.0323 0.0367

to enjoy 0.0214 0.0318 0.0195

rude and 0.0214 0.0141 0.0227

tried to 0.0341 0.0195 0.0367

enjoy 0.0818 0.1113 0.0766

fresh 0.1391 0.2016 0.1282

salad 0.1408 0.194 0.1314

bottle 0.0296 0.0266 0.0301

issues 0.0343 0.0243 0.036

i love 0.1501 0.1556 0.1492

dresses 0.0125 0.0014 0.0144

valet 0.0185 0.0228 0.0178

pasta 0.0613 0.0992 0.0546

issue 0.0397 0.0273 0.0419

lunch 0.1521 0.2338 0.1377

the food 0.2139 1.3898 0.0074

chandler 0.0257 0.0284 0.0252

special 0.1097 0.1363 0.105

the wings 0.028 0.0325 0.0272

appointment 0.045 0.0024 0.0524

worst 0.0589 0.058 0.059

pedicure 0.0285 0.0007 0.0334

dress 0.024 0.0094 0.0266

the office 0.0253 0.0055 0.0288

delicious i 0.0226 0.0448 0.0187

account 0.0148 0.0067 0.0162

i called 0.0394 0.0098 0.0446

returned 0.0213 0.0172 0.022

charge 0.0519 0.033 0.0552

phone 0.0537 0.0251 0.0587

sports 0.0368 0.0406 0.0362

equipment 0.0241 0.0035 0.0277

the other 0.0862 0.0961 0.0844

the manager 0.0544 0.0701 0.0516

stop 0.0812 0.08 0.0814

i went to 0.0388 0.0286 0.0406

me to 0.0628 0.0429 0.0664

view 0.0459 0.0771 0.0404

spicy 0.0677 0.0908 0.0636

chicken and 0.0333 0.06 0.0286

the guy 0.0307 0.0224 0.0321

baked 0.0344 0.0431 0.0328

miles 0.0241 0.0207 0.0247

lack of 0.0241 0.0215 0.0246

the perfect 0.0263 0.0263 0.0263

theater 0.0284 0.0165 0.0305

brunch 0.0215 0.0316 0.0198

zoo 0.0143 0.0072 0.0156

waste 0.0273 0.0205 0.0284

looked like 0.0211 0.0333 0.019

entrees 0.0276 0.0528 0.0232

follow 0.0169 0.0118 0.0177

food at 0.0234 0.0593 0.0171

pain 0.0241 0.0075 0.027

the wrong 0.0165 0.0219 0.0156

hawaiian 0.0224 0.0209 0.0227

votes.cool 0.6009 0.6197 0.5976

experienced 0.0222 0.0282 0.0211

review.rating High Rating High Rating High Rating

Time taken to build model (full training data) : 95.64 seconds

=== Model and evaluation on test split ===

kMeans

======

Number of iterations: 9

Within cluster sum of squared errors: 10978.307684215653

Initial staring points (random):

Cluster 0: {40 5.589311}

Cluster 1: {9 1.65687,51 4.84031,61 4.11174}

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1

(18377) (15626) (2751)

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

airport 0.0352 0.038 0.0194

bruschetta 0.0434 0.0453 0.0331

wine 0.092 0.0859 0.1265

terminal 0.0145 0.0159 0.0066

vegas 0.0757 0.0735 0.088

great 0.3471 0.3515 0.3218

mesa 0.0155 0.0168 0.0085

in vegas 0.0341 0.0336 0.0373

postino 0.0228 0.0207 0.0349

food 0.343 0.2329 0.968

delicious 0.1447 0.1308 0.2234

rental 0.0179 0.0211 0

love 0.2508 0.2527 0.2399

security 0.0163 0.0179 0.0072

henderson 0.0258 0.0245 0.0335

the wine 0.0209 0.0156 0.0507

of wine 0.0267 0.0267 0.0265

place 0.317 0.3019 0.4026

insurance 0.0244 0.0287 0

rude 0.0577 0.0619 0.034

car 0.1076 0.1223 0.0243

customer 0.1228 0.1326 0.067

manager 0.0814 0.0804 0.0871

the airport 0.0142 0.0143 0.0136

the best 0.1866 0.1895 0.1701

office 0.067 0.0762 0.0148

bus 0.0123 0.012 0.0137

greek 0.0213 0.0198 0.0301

called 0.0827 0.0888 0.0479

lgo 0.0165 0.0154 0.0233

customer service 0.0951 0.1048 0.0397

wine and 0.0245 0.0225 0.0358

bottle of 0.0187 0.0186 0.0193

good 0.3271 0.3089 0.4306

italian 0.0746 0.0681 0.111

is great 0.0818 0.0705 0.1456

excellent 0.1258 0.1197 0.161

terrible 0.0487 0.0449 0.07

bread 0.0951 0.0875 0.1381

las vegas 0.0346 0.0339 0.0387

chicken 0.1828 0.1653 0.2822

call 0.0829 0.0877 0.0557

sandwich 0.0896 0.0898 0.0881

wings 0.0881 0.0866 0.0962

grilled 0.0364 0.0337 0.0517

management 0.0334 0.0328 0.0364

told 0.105 0.1103 0.0749

oil change 0.0227 0.0267 0

amazing 0.1524 0.1478 0.1787

company 0.0486 0.0534 0.0211

casino 0.0312 0.0311 0.0317

atmosphere 0.1098 0.0992 0.1705

credit 0.032 0.0345 0.018

cheese 0.126 0.1235 0.1404

las 0.0383 0.0372 0.0449

tomato 0.0342 0.0332 0.0396

deli 0.0376 0.0375 0.0382

to enjoy 0.0214 0.0199 0.03

rude and 0.0196 0.0207 0.0136

tried to 0.0328 0.0345 0.0228

enjoy 0.086 0.0799 0.1207

fresh 0.14 0.1283 0.2069

salad 0.1379 0.1281 0.1939

bottle 0.0272 0.027 0.028

issues 0.0356 0.0372 0.0261

i love 0.145 0.1422 0.1611

dresses 0.0122 0.014 0.0021

valet 0.0182 0.0176 0.0215

pasta 0.0616 0.0552 0.0979

issue 0.0395 0.0418 0.0266

lunch 0.1513 0.1378 0.2282

the food 0.2116 0.0061 1.3793

chandler 0.0253 0.0242 0.0318

special 0.1098 0.1038 0.1443

the wings 0.0261 0.0249 0.0333

appointment 0.0481 0.0559 0.0036

worst 0.0571 0.0575 0.0547

pedicure 0.0303 0.0355 0.0011

dress 0.0259 0.0288 0.0091

the office 0.0248 0.0284 0.0049

delicious i 0.0199 0.0171 0.0358

account 0.0159 0.0174 0.0074

i called 0.0389 0.0435 0.0131

returned 0.0222 0.023 0.0174

charge 0.0493 0.0522 0.0324

phone 0.0523 0.0571 0.025

sports 0.0358 0.0351 0.0392

equipment 0.027 0.031 0.0044

the other 0.0891 0.0873 0.0997

the manager 0.0524 0.0488 0.0723

stop 0.0833 0.0832 0.0837

i went to 0.0376 0.0408 0.0197

me to 0.0647 0.0675 0.0485

view 0.0478 0.0423 0.0794

spicy 0.067 0.0626 0.0915

chicken and 0.0328 0.0271 0.0655

the guy 0.0326 0.0341 0.0243

baked 0.0327 0.0316 0.0386

miles 0.0258 0.0267 0.021

lack of 0.0224 0.0232 0.0179

the perfect 0.0265 0.0263 0.0273

theater 0.0277 0.0294 0.0178

brunch 0.0227 0.0217 0.0283

zoo 0.0145 0.0161 0.0056

waste 0.0269 0.0282 0.0195

looked like 0.0224 0.0196 0.0384

entrees 0.0279 0.0231 0.0553

follow 0.0154 0.016 0.0121

food at 0.0228 0.0159 0.062

pain 0.0252 0.0281 0.0089

the wrong 0.0174 0.0161 0.0248

hawaiian 0.0235 0.0239 0.0213

votes.cool 0.5998 0.5969 0.6161

experienced 0.023 0.0223 0.0271

review.rating High Rating High Rating High Rating

Time taken to build model (percentage split) : 40.66 seconds

Clustered Instances

0 8045 ( 85%)

1 1422 ( 15%)