=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayesMultinomial

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: 5-fold cross-validation

=== Classifier model (full training set) ===

The independent probability of a class

--------------------------------------

High Rating 0.6266968325791855

Low Rating 0.3733031674208145

The probability of a word given the class

-----------------------------------------

High Rating Low Rating

airport 0.001299388898688305 0.01060415351577254

bruschetta 0.008075963031865566 9.642269000358322E-4

wine 0.014122726931867603 0.006478360564247007

terminal 2.503627412234072E-4 0.004579344470420648

vegas 0.01096294543334549 0.005147018760661997

great 0.0449658379671817 0.039906688584391156

mesa 6.953903388158178E-4 0.004574728673144189

in vegas 0.00531351667725509 0.0015927951297412455

postino 0.0039017054134563235 7.639323578079527E-4

food 0.0432668877978695 0.04115382311027545

delicious 0.019694126850762637 0.015228521286292717

rental 9.539794244978912E-4 0.004587340732184718

love 0.032651423297053984 0.02901143279416226

security 0.0012365727868599227 0.0044931763478060095

henderson 0.004023512038409889 0.0013967115627296243

the wine 0.0033784179175212815 0.0010181760058219776

of wine 0.004509155416336014 0.0016272467627252891

place 0.040040737065146846 0.03702566229788549

insurance 0.0016446867671604 0.005089532906517837

rude 0.005926627491345094 0.010390302974717174

car 0.011401152497663731 0.017684376662458246

customer 0.012979517225797708 0.01909633281700407

manager 0.008642173361625304 0.0137974739353817

the airport 9.324575599681361E-4 0.0034702844022432918

the best 0.024335951237179246 0.02059076766090219

office 0.006412551527576072 0.011471232439231257

bus 7.324507263146306E-4 0.0029273294504369145

greek 0.0037140505840375397 0.0012116292832491482

called 0.008686823277154825 0.012857400929370589

lgo 0.0027688638253177264 5.322354805762375E-4

customer service 0.010029838410193358 0.01514698929757804

wine and 0.003684318794373309 0.0015986389262968655

bottle of 0.0030957095229799413 0.0012016032908808494

good 0.04132016212150156 0.04007063020421483

italian 0.010425999883922147 0.006933138677345485

is great 0.011497878078840177 0.008579456004007014

excellent 0.017252555069836858 0.013214552107595092

terrible 0.005193374980080569 0.00917535026812808

bread 0.012585554003309681 0.009622387683031303

las vegas 0.004832514790310347 0.0025921164602921223

chicken 0.024405382799938792 0.02035277949512094

call 0.008671783016991093 0.012892155109686393

sandwich 0.012759676111568884 0.009479803659173896

wings 0.01322529278717699 0.008806266828723602

grilled 0.005574015388486011 0.0031315290804129424

management 0.003108815957781443 0.0061948834759643565

told 0.011322634091875186 0.015711894836641854

oil change 0.0019708106306291946 0.004638726437389929

amazing 0.01950431573213644 0.01638866490675076

company 0.004583220364246977 0.008031546064694403

casino 0.004893669068557803 0.002415663158071537

atmosphere 0.014511503593035147 0.011673387499030491

credit 0.003105645816838435 0.00622110587575544

cheese 0.01672485766322845 0.01416600909941625

las 0.005292566138754439 0.00304371291480437

tomato 0.005116619824196159 0.0031296304409164084

deli 0.00529788719039939 0.0029719640375471075

to enjoy 0.003288333385362328 0.0014986659955099632

rude and 0.0018382841459487187 0.004089384454675569

tried to 0.0033446116088765576 0.005788097883949875

enjoy 0.010974516114486263 0.008546225283296967

fresh 0.018271122287237977 0.015243776946162187

salad 0.01836950080704106 0.01562957526596202

bottle 0.004328898689993864 0.0024741136830334076

issues 0.0032941729747706745 0.005921440798179599

i love 0.019514455744675738 0.01679617078499445

dresses 8.133500448644563E-4 0.002858974595416269

valet 0.00288444612120966 0.0012476804717865722

pasta 0.008408396311382875 0.006082548491028661

issue 0.003969036036467238 0.006599010837106196

lunch 0.019773037382131514 0.017008666746695784

the food 0.026869990345848305 0.02559733732411814

chandler 0.0022893465304717063 0.004762680659731397

special 0.014378508140060293 0.01206867236069077

the wings 0.004240620325028234 0.0020740453640423332

appointment 0.004460157667915319 0.007514466569881952

worst 0.005913950610153358 0.009693063390503097

pedicure 0.0025557615007714605 0.005256277553973539

dress 0.001990654359930271 0.004733486586235802

the office 0.0023308314966657695 0.004572028659754148

delicious i 0.003314738642097325 0.001872867287961948

account 0.0010939979803518882 0.003158037671960393

i called 0.00404288661251871 0.006334669981565885

returned 0.0019324231159501138 0.003889538476736286

charge 0.005438040513463582 0.00815308498637504

phone 0.0055091280467485584 0.008645448408937504

sports 0.005367664839563351 0.003105518847068633

equipment 0.0021975693681157544 0.004384935889347807

the other 0.00934821350807665 0.012955259738728284

the manager 0.005678119489995371 0.008578741066926052

stop 0.010638210179534814 0.008940343969013567

i went to 0.003940890435200667 0.006310088044754517

me to 0.00681496236775926 0.00945896262577016

view 0.006509478771556711 0.004193680742092795

spicy 0.00919446454849138 0.0068843645643563035

chicken and 0.004701782912843689 0.003084745368230647

the guy 0.0029810943889662877 0.005241648607214659

baked 0.004877721052350599 0.003136970482977957

miles 0.0023461216462068667 0.004123333489578597

lack of 0.002438348775573524 0.003959864677763839

the perfect 0.003724249123299402 0.0024358654982446514

theater 0.004184498341213079 0.0023221146775798843

brunch 0.00322317223131986 0.00167643296965822

zoo 0.002350623669842673 7.625945037861348E-4

waste 0.0026056623812438216 0.004740805342631404

looked like 0.002237222423665123 0.0032830898870932473

entrees 0.0038778894398437493 0.0025876569468860624

follow 0.0015616199316865716 0.0030316030840880684

food at 0.0033594037724980663 0.0020712729645978077

pain 0.00255320851736912 0.003748429683873842

the wrong 0.0015116013747761177 0.003009661937954017

hawaiian 0.0032852663428233784 0.0018610122073495338

votes.cool 0.07314628949161074 0.07599734581214622

experienced 0.0024048951916687752 0.0033467942123495166

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 17255 61.9703 %

Incorrectly Classified Instances 10589 38.0297 %

Kappa statistic 0.1415

Mean absolute error 0.4137

Root mean squared error 0.5255

Relative absolute error 88.4078 %

Root relative squared error 108.6558 %

Coverage of cases (0.95 level) 92.4831 %

Mean rel. region size (0.95 level) 87.5269 %

Total Number of Instances 27844

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.775 0.641 0.670 0.775 0.719 0.145 0.612 0.722 High Rating

0.359 0.225 0.487 0.359 0.413 0.145 0.612 0.491 Low Rating

Weighted Avg. 0.620 0.486 0.602 0.620 0.605 0.145 0.612 0.636

=== Confusion Matrix ===

a b <-- classified as

13523 3927 | a = High Rating

6662 3732 | b = Low Rating