=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayesMultinomial

Relation: yelp\_task2\_trainingset\_binaryclassification-weka.filters.unsupervised.attribute.NominalToString-Cfirst-weka.filters.unsupervised.attribute.StringToWordVector-R1-W2000-prune-rate-1.0-C-T-I-N1-L-S-stemmerweka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.NGramTokenizer -max 1 -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: 142

[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 0.0014324443943396726 0.010939653663914322

bruschetta 0.008550204512976388 0.0010726431906898906

wine 0.015547388763103196 0.007072537209259656

terminal 2.498801755515264E-4 0.0044791429224504365

vegas 0.012360116421060156 0.005795980121506248

great 0.05234797969052783 0.04601241975629845

mesa 7.531118666036958E-4 0.004964816812700306

terminals 2.9676076542396233E-5 0.0019218397782349316

postino 0.004079814945963587 8.481361824739372E-4

harbor 1.940690922819912E-4 0.0023915683079013756

food 0.05122003438135196 0.04854111682684738

delicious 0.021770710652817656 0.017392762325348962

rental 0.0010303483204966862 0.004604026179782445

love 0.03754947718806619 0.03339610842476859

place 0.04687152293167572 0.043674009025045285

henderson 0.004303714253415002 0.0016212145204180883

postinos 0.0023645212347227383 9.592781556388297E-5

fig 0.0020812531077759188 1.8833727757120922E-4

security 0.0013151762529545405 0.004557127780632208

insurance 0.0017192143913087902 0.005292081087063031

car 0.01255898864780681 0.019323420328915063

rude 0.006510983045878296 0.011493409036594704

gate 4.559770642808585E-4 0.0027459688501581957

customer 0.014712041564444792 0.021131204448828902

sky 6.105411254914934E-4 0.002792843560139417

manager 0.009662656679644664 0.01535263406444269

panini 0.0019639485355716526 1.8741124109396222E-4

lgo 0.0028372220495039187 5.672728612659193E-4

called 0.009595097486920562 0.014171601098819963

flight 5.767621161275638E-4 0.0024825750692118384

goat 0.0021297576121588526 5.060791221192619E-4

casa 3.6966341991863735E-4 0.002070030196398901

office 0.006923462064527964 0.012285181890297406

call 0.009523692240843696 0.014223839715624602

bus 7.837849026728538E-4 0.003034812013571295

wickenburg 0.0019523275252740383 2.4635281482262146E-4

greek 0.003950903669862827 0.0013778595338278642

italian 0.011422776303997323 0.007767647653351583

pho 6.923836734615111E-4 0.003569996379641876

good 0.04796239660793825 0.046449505260163504

chicken 0.026693309044243423 0.022648383780889073

bread 0.01350362943064749 0.010532820384831734

excellent 0.0190054352757433 0.015291767484313436

wings 0.014781956824618385 0.00966484209064141

management 0.0033192666441566117 0.006716757591030251

brie 0.0022369066643565722 6.952964842291376E-4

sandwich 0.013703070927984179 0.01063794801478567

casino 0.0052205445020371884 0.002541399006141428

prosciutto 0.00210877097224638 5.582549034480155E-4

pesto 0.0026620528142299676 9.722415933941695E-4

amazing 0.021977664607848384 0.018838928997822704

told 0.012779120005166159 0.01778595313938979

dress 0.0021476831784103304 0.00523163364565884

company 0.004945070892006076 0.008266901835915478

las 0.00589994067382759 0.0033437474103706474

atmosphere 0.01658998961871316 0.013361411352312091

terrible 0.005975951208436879 0.010107644579365948

cheese 0.017819701547084114 0.015575361514448717

tomato 0.005312352241161876 0.0033468239381455028

deli 0.0054384829482466285 0.0031470391462123663

grilled 0.005881941537162082 0.0033908428216624462

fresh 0.01991434942272466 0.01676282649296956

arcadia 0.0026011289546705235 7.042093762180701E-4

shuttle 7.595951492360406E-4 0.002464958475439942

teds 0.0024912853467440026 5.677674175628531E-4

credit 0.003332171872239362 0.006609704692678731

bottle 0.004666687208181943 0.0026785281170105613

returned 0.002081320930747191 0.00416498264591767

royal 0.0019232344078995928 5.579176133511162E-4

dresses 8.549107075463804E-4 0.003200547917977093

madison 0.0022168188629161714 4.691059810953379E-4

chandler 0.002476576196243663 0.005127753522108399

sports 0.005689019336183857 0.0033051033798105313

issues 0.0036581573622846354 0.0064040842686894665

salad 0.020040388087078993 0.017171083105946124

enjoy 0.01222396791211887 0.009380459550380114

lunch 0.022314300516066483 0.01940022677953573

issue 0.004351456237155927 0.0072609151927056

trader 0.002570242789567062 7.918239388830348E-4

valet 0.00301241508880689 0.0013152731534149496

pedicure 0.002816810257919159 0.005768213575722272

billing 3.5692632911830816E-4 0.0015583681286974304

pasta 0.009208575495467023 0.006687982805793269

appointment 0.004967548444976573 0.008492694915970177

companies 6.270823474595366E-4 0.0020416583045064036

yard 8.652049767778109E-4 0.002070227253744684

charge 0.005810698982277785 0.008758717275603166

worst 0.006522227325830619 0.010863761704786652

account 0.0011725492135432987 0.0034292732118324475

equipment 0.002357692308488453 0.004601567493935162

fees 0.0011686046719482802 0.002993182303884026

guy 0.008186797559229731 0.012110164565646469

emergency 0.0010087269272844104 0.0021500181952382453

phone 0.0059907836841005865 0.009496992395247587

change 0.006819968997593946 0.010390085411916059

lexus 1.5360330113261376E-4 0.0015024405574550507

mall 0.009421699625329176 0.008166448489978039

cell 9.278203701726277E-4 0.0020425667249504755

perfect 0.012968524414872786 0.010623728305654896

jewish 0.0017176353421670357 4.894398626667395E-4

stop 0.011743874861414153 0.009921867594961258

view 0.007075297214388279 0.00452857077732156

korean 0.0030940474414002522 0.001356746177618885

awful 0.0037464319805728687 0.006565082455049363

weiss 0.0015550299654231757 3.7399731954872715E-4

dipping 0.0016418559918034668 8.937781658262476E-4

claim 8.507519287691182E-4 0.0024194531351601407

posted 0.0011110028085968644 0.0025618769658001727

bridal 3.7806833500886355E-4 0.0014976071872274924

joes 0.00298551914824259 0.0014910358588066531

unit 6.571020459390979E-4 0.002288627606911431

zoo 0.0024067002445389302 8.047789803927406E-4

baked 0.005069806675120474 0.0032894916589498837

ross 0.0026426584605075583 0.001073081968373195

police 4.5154926759689206E-4 0.0015773676771764103

warranty 0.0013448495511039185 0.002930037311624459

palms 0.0014489252081367782 4.5742860897200364E-4

rep 6.333309214965148E-4 0.0017830508708815964

theater 0.004228842212429567 0.0024230465311531244

spanish 9.98899897675979E-4 0.0020360884437272656

chicago 0.005509826522242002 0.0035404175943977243

follow 0.001678144089986872 0.0032016953020005496

travel 0.0016196420972244409 0.0030969267603369165

smoked 0.0019358780581753198 9.520421158601409E-4

entree 0.004665518128186544 0.0035644882423167394

garbage 8.175437892231012E-4 0.0024263295546843528

romantic 0.0028169424872676457 0.0015349937880531428

cozy 0.0031939112542756385 0.0016199122653518052

rented 6.297087793851856E-4 0.0016599611301208396

brunch 0.003415140521234726 0.001784787126121787

votes.cool 0.07982964500570032 0.08359452234841415

hawaiian 0.0034782840753818304 0.0019412327832817712

pastrami 0.002186068826888106 9.254515598361336E-4

breads 0.001974284376759326 9.939607501771432E-4

sliders 0.002378884106697577 0.0011310724552830867

village 0.0016744062385290071 6.229375643812741E-4

pub 0.0018873186845029567 7.721118711960397E-4

popcorn 0.0019693535173606405 8.132330477173434E-4

installed 6.581247935911629E-4 0.001667940144223284

win 0.002972027518153156 0.0013685748656659688

hillstone 0.0014284294567825106 6.4249626740317E-4

Time taken to build model: 0.08 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 17457 62.6957 %

Incorrectly Classified Instances 10387 37.3043 %

Kappa statistic 0.1555

Mean absolute error 0.408

Root mean squared error 0.5181

Relative absolute error 87.1933 %

Root relative squared error 107.1238 %

Coverage of cases (0.95 level) 92.8638 %

Mean rel. region size (0.95 level) 87.1714 %

Total Number of Instances 27844

=== Detailed Accuracy By Class ===

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

0.785 0.638 0.674 0.785 0.725 0.160 0.624 0.733 High Rating

0.362 0.215 0.500 0.362 0.420 0.160 0.624 0.503 Low Rating

Weighted Avg. 0.627 0.480 0.609 0.627 0.611 0.160 0.624 0.647

=== Confusion Matrix ===

a b <-- classified as

13693 3757 | a = High Rating

6630 3764 | b = Low Rating