=== 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 3 -min 3 -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: 54

spicy korean chicken

the wine list

should have been

in las vegas

at the front

bottle of wine

to try the

was really good

glass of wine

wife had the

the royal palms

a phone call

have a lot

feel like i

taken care of

we would have

i took it

next to the

i look forward

an oil change

customer service i

cant go wrong

the place and

what the hell

for an oil

i told them

of wine and

a bottle of

la grande orange

of the day

votes.cool

look forward to

at least the

this is not

on the same

hung up on

food is delicious

i called to

great food and

for a place

dont have to

chicken noodle soup

right next to

to have it

as if they

for my taste

at the time

a few more

never come back

i asked her

last time we

have gone to

a gift card

review.rating

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

spicy korean chicken 0.009530960083681136 8.796543672590768E-4

the wine list 0.008809263618518344 0.001786344868162385

should have been 0.005660692896911351 0.015344798135569472

in las vegas 0.01703370108104166 0.00658912949836123

at the front 0.009588890016355264 0.0195553580469899

bottle of wine 0.014250876438265093 0.0051928700208388814

to try the 0.018934909137549007 0.009838461148649974

was really good 0.020035126593078156 0.009691278900972553

glass of wine 0.00950679357490512 0.004323210576549032

wife had the 0.008159452037256282 0.0019071567217050652

the royal palms 0.006591326211832732 0.0014256830430639561

a phone call 0.0016065073172830582 0.008255433017180545

have a lot 0.012985183808402 0.004869379264386789

feel like i 0.011654215466167874 0.0064675756932827035

taken care of 0.007372136448191537 0.014857381518499625

we would have 0.005136681438867095 0.013072579430838237

i took it 0.002520868694881376 0.010232811989775076

next to the 0.011536269642548891 0.007961749034763321

i look forward 0.00799130233390872 0.0028497934466332874

an oil change 0.00757293045746037 0.01763549212587051

customer service i 0.007110345462179977 0.017594030514146437

cant go wrong 0.020572215151630834 0.010884670380250048

the place and 0.007681920336109269 0.0042807387458863415

what the hell 0.0016823399420871946 0.00746164317823045

for an oil 0.004296749571772988 0.011288988877648563

i told them 0.003791544663391716 0.011562163586676295

of wine and 0.010870571293068211 0.004333942429511178

a bottle of 0.01284048312277885 0.006047083430952687

la grande orange 0.006108693393164992 6.991252838417384E-4

of the day 0.017962286266287517 0.009338375421721464

votes.cool 0.5351299212431878 0.5691788691721462

look forward to 0.013402264517210348 0.006365349441543677

at least the 0.00326870659190622 0.009624276524180044

this is not 0.017206163912630457 0.025481534369593523

on the same 0.003177715505504868 0.0080101984282058

hung up on 0.002317351145432473 0.00857116375726157

food is delicious 0.009132005862439261 0.0032263356891765595

i called to 0.004044876299236142 0.010589077327918573

great food and 0.02198771880784909 0.011973893476757713

for a place 0.008904780842321632 0.0049405588412236665

dont have to 0.014148485095735899 0.0076581739820616515

chicken noodle soup 0.007842512951043998 0.0014360263017247645

right next to 0.00916833905344463 0.005516967298283234

to have it 0.004240280126703718 0.010879752820791736

as if they 0.0026873374928200163 0.008286859665687455

for my taste 0.008855058307862202 0.003683980229605298

at the time 0.011674194904282896 0.012552094008175176

a few more 0.00947113037175134 0.004228483349267679

never come back 0.003114055187566353 0.009421230626572215

i asked her 0.004364990097937513 0.008275644468366872

last time we 0.007935605145201021 0.0050006798405050675

have gone to 0.004894103853570865 0.009394980343949828

a gift card 0.0036371661847854615 0.009476967338784469

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 17542 63.001 %

Incorrectly Classified Instances 10302 36.999 %

Kappa statistic 0.0445

Mean absolute error 0.4536

Root mean squared error 0.4876

Relative absolute error 96.9414 %

Root relative squared error 100.8159 %

Coverage of cases (0.95 level) 98.7538 %

Mean rel. region size (0.95 level) 97.3046 %

Total Number of Instances 27844

=== Detailed Accuracy By Class ===

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

0.959 0.922 0.636 0.959 0.765 0.078 0.537 0.658 High Rating

0.078 0.041 0.530 0.078 0.136 0.078 0.537 0.413 Low Rating

Weighted Avg. 0.630 0.593 0.596 0.630 0.530 0.078 0.537 0.566

=== Confusion Matrix ===

a b <-- classified as

16730 720 | a = High Rating

9582 812 | b = Low Rating