=== 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 2 -min 2 -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: 43

in vegas

the bruschetta

of wine

the wine

the airport

the best

customer service

wine and

bottle of

is great

oil change

las vegas

tried to

to enjoy

rude and

money and

i love

do not

this office

in henderson

the casino

the wings

called to

the food

the office

i called

the other

the manager

me to

hot dog

a bottle

chicken and

the guy

lack of

delicious i

place and

the perfect

told them

the wrong

but nothing

votes.cool

glass of

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

in vegas 0.0184975364370358 0.0054441837864484515

the bruschetta 0.011620059440810857 0.0014752568181943938

of wine 0.015934310498628826 0.005628855712742692

the wine 0.011773996262208007 0.00351392264550666

the airport 0.0032687780020917933 0.012687643439275057

the best 0.08718398887871831 0.07327137105756634

customer service 0.03543265338281792 0.0545087938028196

wine and 0.013192441603667716 0.005700423200227022

bottle of 0.011001369832826142 0.0041908875988972685

is great 0.04024113044148108 0.029584490003176832

oil change 0.006732209061926904 0.015885654034263307

las vegas 0.01730706169908628 0.009231446435121539

tried to 0.011398404975706454 0.02012051940261635

to enjoy 0.011555634425696308 0.005184263400304653

rude and 0.006329625180836364 0.014192896523148684

money and 0.003604848822225951 0.009547165198817228

i love 0.06839743997829019 0.05910160419707288

do not 0.023478473488961826 0.03527268580921776

this office 0.0023664861365778576 0.007726334536467478

in henderson 0.00799825114642689 0.003187728273550774

the casino 0.007656683307616918 0.0027389973504496072

the wings 0.014523452192864874 0.0071968285773722105

called to 0.003975943212498875 0.009758174804852351

the food 0.0925783601657143 0.0882484884372464

the office 0.008096543140011294 0.015848909367221855

i called 0.0138759366848611 0.02144768634687157

the other 0.03239637582235391 0.04443547364745102

the manager 0.019133249644015055 0.028909866181727327

me to 0.02325806853158359 0.03254859637606395

hot dog 0.0091515748202731 0.003781769175097035

a bottle 0.00798558013613137 0.0036518903440116916

chicken and 0.016634469504150817 0.010803138858135743

the guy 0.010119151995058947 0.01784802239647709

lack of 0.008647975070202594 0.014116370723742448

delicious i 0.011507201083393225 0.006335221274564135

place and 0.019103698795950107 0.014795987622139928

the perfect 0.013279816978557676 0.008527257216762694

told them 0.0032996232472003867 0.0074369784113278185

the wrong 0.00523651760435105 0.010430478931556861

but nothing 0.009424409742896146 0.005832183588667897

votes.cool 0.2535528451444606 0.26395234015115343

glass of 0.009247823479833809 0.005899214341667884

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 17221 61.8482 %

Incorrectly Classified Instances 10623 38.1518 %

Kappa statistic 0.0784

Mean absolute error 0.4469

Root mean squared error 0.4934

Relative absolute error 95.5207 %

Root relative squared error 102.016 %

Coverage of cases (0.95 level) 98.2977 %

Mean rel. region size (0.95 level) 96.4553 %

Total Number of Instances 27844

=== Detailed Accuracy By Class ===

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

0.865 0.796 0.646 0.865 0.740 0.091 0.567 0.687 High Rating

0.204 0.135 0.474 0.204 0.285 0.091 0.567 0.437 Low Rating

Weighted Avg. 0.618 0.549 0.582 0.618 0.570 0.091 0.567 0.594

=== Confusion Matrix ===

a b <-- classified as

15100 2350 | a = High Rating

8273 2121 | b = Low Rating