Pilot Study Proposal (~600 words)

Email marketing campaigns are an important way of reaching out to potential customers and it is very likely that machine learning techniques can be helpful in increasing the effectiveness of such campaigns. Given that there is data available about the success/failure of previous emails sent to customers, this is a supervised learning task in which a classifier can be trained using the result of previous campaigns and their outcomes.

Choosing the right information on which to train our classifier is likely to have a big impact on its effectiveness. We would need as much relevant information about each email as possible from the following:

* Recipient features (personal customers): Age, gender, occupation, purchase history, marital status etc.
* Recipient features (business customers): Type of business, turnover, staff strength, purchase history etc.
* Product features: Type, price, size, discount on usual price, colour etc.
* Email features: Words, style, design

Given that this is a supervised classification problem, we could potentially use the following algorithms:

* Decision Tree: A big advantage of decision trees is that rules generated by decision trees are easily comprehensible, which makes them a very useful tool for exploration of a problem. A potential problem with decision trees is that they can over-fit the data, hence care needs to be taken by employing appropriate pruning methods where needed.
* Naïve Bayes: Although Naïve Bayes make a strong assumption of conditional independence in attributes given the classification variable, they can be very powerful in certain situations where such an assumption is justified. The main advantage of Naïve Bayes is their computational efficiency and the fact that they need much less data then would be needed for a complete Bayes classifier.
* K-nearest neighbour (KNN): KNN can be used for the given situation as it assigns patterns to the majority class amongst K nearest neighbours. KNN’s advantage is its simplicity in application and the algorithm often gives good results for linearly separable problems but struggle with non-linearly separable situations.
* Support Vector Machine (SVM): SVMs can overcome a major limitation of some other techniques such as KNN and Decision Trees that these techniques do not give good results for problems which are not linearly separable. SVMs can efficiently classify non-linearly separable data with use of appropriate kernels such as a polynomial or Gaussian kernel. However the performance of SVMs comes at the cost of being more computationally expensive.
* A combination of multiple models by using bagging, boosting or stacking. Bagging and boosting combine outputs from same classifiers, whereas stacking uses metalearning to choose between the outputs of several learning systems. The advantage of combining outputs from multiple models is getting a more robust as well as more accurate classifier. The disadvantage of using multiple models is the increased complexity of the final model.

For the given situation and considering the commercial importance of improving the effectiveness of your email marketing campaign, we would recommend exploring a multiple model combination combining outputs from various algorithms, so that we have a higher chance of providing a better predictions.

The performance of the system in terms of accuracy on unseen data will need to be judged in the context of the frequency of the modal class. The modal class frequency represents the baseline performance for our system which we need to improve upon. In addition to this, we can make use of Kappa statistic which represents an improvement that is achieved over prediction by random guessing.

Appendix A

J48 with minimum 10 instances per leaf

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 10

Relation: ce802ass\_train\_norm

Instances: 1000

Attributes: 9

V1

V2

V3

V4

V5

V6

V7

V8

Response

Test mode: 10-fold cross-validation

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

J48 pruned tree

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

V4 <= 0.658933

| V3 <= 0.589744: FALSE (690.0/230.0)

| V3 > 0.589744

| | V2 <= 0.481651: TRUE (37.0/4.0)

| | V2 > 0.481651

| | | V1 <= 0.571429

| | | | V2 <= 0.830275: TRUE (26.0/5.0)

| | | | V2 > 0.830275: FALSE (15.0/5.0)

| | | V1 > 0.571429: FALSE (53.0/16.0)

V4 > 0.658933

| V3 <= 0.5

| | V6 <= 0.538462: FALSE (14.0/3.0)

| | V6 > 0.538462: TRUE (10.0/3.0)

| V3 > 0.5: TRUE (155.0/6.0)

Number of Leaves : 8

Size of the tree : 15

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 684 68.4 %

Incorrectly Classified Instances 316 31.6 %

Kappa statistic 0.3401

Mean absolute error 0.397

Root mean squared error 0.4545

Relative absolute error 79.8047 %

Root relative squared error 91.1435 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

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

0.946 0.619 0.639 0.946 0.762 0.404 0.680 0.660 FALSE

0.381 0.054 0.859 0.381 0.528 0.404 0.680 0.693 TRUE

Weighted Avg. 0.684 0.357 0.741 0.684 0.654 0.404 0.680 0.675

=== Confusion Matrix ===

a b <-- classified as

507 29 | a = FALSE

287 177 | b = TRUE

Appendix B

KNN with 5 neighbours and stratified cross validation

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 5 -W 0 -X -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: ce802ass\_train\_norm

Instances: 1000

Attributes: 9

V1

V2

V3

V4

V5

V6

V7

V8

Response

Test mode: 10-fold cross-validation

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

IB1 instance-based classifier

using 5 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 782 78.2 %

Incorrectly Classified Instances 218 21.8 %

Kappa statistic 0.5607

Mean absolute error 0.2858

Root mean squared error 0.3905

Relative absolute error 57.4503 %

Root relative squared error 78.3049 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

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

0.812 0.252 0.788 0.812 0.800 0.561 0.855 0.841 FALSE

0.748 0.188 0.775 0.748 0.761 0.561 0.855 0.824 TRUE

Weighted Avg. 0.782 0.223 0.782 0.782 0.782 0.561 0.855 0.833

=== Confusion Matrix ===

a b <-- classified as

435 101 | a = FALSE

117 347 | b = TRUE

Appendix C

SVM with Quadratic Kernel, C=5.

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 5.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 2.0 -L -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

Relation: ce802ass\_train\_norm

Instances: 1000

Attributes: 9

V1

V2

V3

V4

V5

V6

V7

V8

Response

Test mode: 10-fold cross-validation

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

SMO

Kernel used:

Poly Kernel with lower order: K(x,y) = (<x,y> + 1)^2.0

Classifier for classes: FALSE, TRUE

Number of support vectors: 601

Number of kernel evaluations: 497561 (96.699% cached)

Time taken to build model: 0.72 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 724 72.4 %

Incorrectly Classified Instances 276 27.6 %

Kappa statistic 0.4308

Mean absolute error 0.276

Root mean squared error 0.5254

Relative absolute error 55.4864 %

Root relative squared error 105.3436 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

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

0.905 0.485 0.683 0.905 0.778 0.462 0.710 0.669 FALSE

0.515 0.095 0.824 0.515 0.634 0.462 0.710 0.650 TRUE

Weighted Avg. 0.724 0.304 0.749 0.724 0.711 0.462 0.710 0.660

=== Confusion Matrix ===

a b <-- classified as

485 51 | a = FALSE

225 239 | b = TRUE

Appendix D

Stacking model with J48 as the level 1 learner

Level 0 learners:

SVM – Linear C:1

Naïve Bayes

KNN with 1 neighbour

J48 Pruned

=== Run information ===

Scheme: weka.classifiers.meta.Stacking -X 10 -M "weka.classifiers.trees.J48 -C 0.25 -M 2" -S 1 -num-slots 1 -B "weka.classifiers.lazy.IBk -K 1 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.EuclideanDistance -R first-last\\\"\"" -B "weka.classifiers.trees.J48 -C 0.25 -M 2" -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007\" -calibrator \"weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4\""

Relation: ce802ass\_train\_norm

Instances: 1000

Attributes: 9

V1

V2

V3

V4

V5

V6

V7

V8

Response

Test mode: 10-fold cross-validation

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

Stacking

Base classifiers

IB1 instance-based classifier

using 1 nearest neighbour(s) for classification

J48 pruned tree

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

V4 <= 0.658933

| V4 <= 0.24478

| | V4 <= 0.241299

| | | V4 <= 0.237819: FALSE (245.0/26.0)

| | | V4 > 0.237819

| | | | V4 <= 0.238979: TRUE (3.0)

| | | | V4 > 0.238979

| | | | | V6 <= 0.358974: FALSE (2.0)

| | | | | V6 > 0.358974: TRUE (4.0)

| | V4 > 0.241299: FALSE (12.0)

| V4 > 0.24478

| | V3 <= 0.589744: FALSE (442.0/201.0)

| | V3 > 0.589744

| | | V2 <= 0.481651

| | | | V4 <= 0.607889

| | | | | V1 <= 0.928571

| | | | | | V4 <= 0.602088: TRUE (12.0)

| | | | | | V4 > 0.602088: FALSE (2.0)

| | | | | V1 > 0.928571: FALSE (2.0)

| | | | V4 > 0.607889: TRUE (21.0)

| | | V2 > 0.481651

| | | | V1 <= 0.571429: TRUE (37.0/12.0)

| | | | V1 > 0.571429: FALSE (39.0/13.0)

V4 > 0.658933

| V3 <= 0.5

| | V6 <= 0.794872

| | | V6 <= 0.179487: TRUE (4.0/1.0)

| | | V6 > 0.179487: FALSE (14.0/1.0)

| | V6 > 0.794872: TRUE (6.0)

| V3 > 0.5

| | V3 <= 0.75641

| | | V1 <= 0.761905

| | | | V2 <= 0.444954: TRUE (42.0)

| | | | V2 > 0.444954

| | | | | V2 <= 0.605505: TRUE (7.0/1.0)

| | | | | V2 > 0.605505: FALSE (2.0)

| | | V1 > 0.761905

| | | | V6 <= 0.367521: TRUE (2.0)

| | | | V6 > 0.367521: FALSE (3.0)

| | V3 > 0.75641: TRUE (99.0)

Number of Leaves : 21

Size of the tree : 41

Naive Bayes Classifier

Class

Attribute FALSE TRUE

(0.54) (0.46)

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

V1

mean 0.547 0.4799

std. dev. 0.2869 0.2679

weight sum 536 464

precision 0.0238 0.0238

V2

mean 0.5572 0.3761

std. dev. 0.2734 0.2354

weight sum 536 464

precision 0.0047 0.0047

V3

mean 0.3009 0.5104

std. dev. 0.2213 0.3058

weight sum 536 464

precision 0.0128 0.0128

V4

mean 0.322 0.5295

std. dev. 0.1942 0.205

weight sum 536 464

precision 0.0022 0.0022

V5

mean 0.4591 0.4688

std. dev. 0.255 0.2439

weight sum 536 464

precision 0.0556 0.0556

V6

mean 0.5199 0.5043

std. dev. 0.2464 0.2595

weight sum 536 464

precision 0.0087 0.0087

V7

mean 0.9776 0.9828

std. dev. 0.1667 0.1667

weight sum 536 464

precision 1 1

V8

mean 0.483 0.4647

std. dev. 0.2788 0.2722

weight sum 536 464

precision 0.0094 0.0094

SMO

Kernel used:

Linear Kernel: K(x,y) = <x,y>

Classifier for classes: FALSE, TRUE

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.3833 \* (normalized) V1

+ -1.1829 \* (normalized) V2

+ 1.7267 \* (normalized) V3

+ 1.8753 \* (normalized) V4

+ 0.0833 \* (normalized) V5

+ -0.107 \* (normalized) V6

+ 0.0079 \* (normalized) V7

+ -0.2465 \* (normalized) V8

- 0.7858

Number of kernel evaluations: 84685 (70.356% cached)

Meta classifier

J48 pruned tree

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

weka.classifiers.bayes.NaiveBayes-3:FALSE <= 0.088111: TRUE (188.0/5.0)

weka.classifiers.bayes.NaiveBayes-3:FALSE > 0.088111

| weka.classifiers.bayes.NaiveBayes-3:FALSE <= 0.951173

| | weka.classifiers.lazy.IBk-1:FALSE <= 0.001109: TRUE (271.0/98.0)

| | weka.classifiers.lazy.IBk-1:FALSE > 0.001109: FALSE (413.0/108.0)

| weka.classifiers.bayes.NaiveBayes-3:FALSE > 0.951173: FALSE (128.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0.39 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 790 79 %

Incorrectly Classified Instances 210 21 %

Kappa statistic 0.577

Mean absolute error 0.3045

Root mean squared error 0.39

Relative absolute error 61.2187 %

Root relative squared error 78.2089 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

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

0.817 0.241 0.796 0.817 0.807 0.577 0.842 0.840 FALSE

0.759 0.183 0.782 0.759 0.770 0.577 0.842 0.828 TRUE

Weighted Avg. 0.790 0.214 0.790 0.790 0.790 0.577 0.842 0.835

=== Confusion Matrix ===

a b <-- classified as

438 98 | a = FALSE

112 352 | b = TRUE

Appendix

J48 with Clustering

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 10

Relation: ce802ass\_train\_norm-weka.filters.unsupervised.attribute.AddCluster-Wweka.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

Instances: 1000

Attributes: 10

V1

V2

V3

V4

V5

V6

V7

V8

Response

cluster

Test mode: 10-fold cross-validation

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

J48 pruned tree

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

V4 <= 0.658933

| V4 <= 0.24478: FALSE (266.0/33.0)

| V4 > 0.24478

| | V3 <= 0.589744

| | | cluster = cluster1

| | | | V6 <= 0.74359: TRUE (170.0/75.0)

| | | | V6 > 0.74359

| | | | | V8 <= 0.196262: TRUE (10.0/2.0)

| | | | | V8 > 0.196262: FALSE (46.0/7.0)

| | | cluster = cluster2

| | | | V6 <= 0.91453: FALSE (205.0/80.0)

| | | | V6 > 0.91453: TRUE (11.0)

| | V3 > 0.589744

| | | V2 <= 0.481651: TRUE (37.0/4.0)

| | | V2 > 0.481651

| | | | V1 <= 0.571429: TRUE (37.0/12.0)

| | | | V1 > 0.571429: FALSE (39.0/13.0)

V4 > 0.658933

| V3 <= 0.5

| | V6 <= 0.538462: FALSE (14.0/3.0)

| | V6 > 0.538462: TRUE (10.0/3.0)

| V3 > 0.5: TRUE (155.0/6.0)

Number of Leaves : 12

Size of the tree : 23

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 740 74 %

Incorrectly Classified Instances 260 26 %

Kappa statistic 0.4755

Mean absolute error 0.3082

Root mean squared error 0.4164

Relative absolute error 61.9606 %

Root relative squared error 83.5047 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

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

0.780 0.306 0.746 0.780 0.763 0.476 0.818 0.825 FALSE

0.694 0.220 0.732 0.694 0.712 0.476 0.818 0.804 TRUE

Weighted Avg. 0.740 0.266 0.740 0.740 0.739 0.476 0.818 0.815

=== Confusion Matrix ===

a b <-- classified as

418 118 | a = FALSE

142 322 | b = TRUE

Appendix

KNN with clustering

=== Run information ===

Scheme: weka.classifiers.lazy.IBk -K 5 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Relation: ce802ass\_train\_norm-weka.filters.unsupervised.attribute.AddCluster-Wweka.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

Instances: 1000

Attributes: 10

V1

V2

V3

V4

V5

V6

V7

V8

Response

cluster

Test mode: 10-fold cross-validation

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

IB1 instance-based classifier

using 5 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 784 78.4 %

Incorrectly Classified Instances 216 21.6 %

Kappa statistic 0.5649

Mean absolute error 0.2775

Root mean squared error 0.3835

Relative absolute error 55.7877 %

Root relative squared error 76.8988 %

Total Number of Instances 1000

=== Detailed Accuracy By Class ===

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

0.812 0.248 0.791 0.812 0.801 0.565 0.864 0.842 FALSE

0.752 0.188 0.776 0.752 0.764 0.565 0.864 0.828 TRUE

Weighted Avg. 0.784 0.220 0.784 0.784 0.784 0.565 0.864 0.836

=== Confusion Matrix ===

a b <-- classified as

435 101 | a = FALSE

115 349 | b = TRUE