=== Run information ===

Scheme: weka.classifiers.trees.Id3

Relation: input-weka.filters.unsupervised.attribute.Remove-R1

Instances: 20

Attributes: 6

Income

Gender

Age

Holiday

Wine

Take-up

Test mode: 10-fold cross-validation

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

Id3

Gender = Male

| Wine = Yes

| | Income = 40-50K

| | | Age = 30-40: Yes

| | | Age = 20-30: null

| | | Age = 40-50: No

| | Income = 30-40K: Yes

| | Income = 50-60K: Yes

| | Income = 20-30K: Yes

| Wine = No: No

Gender = Female

| Income = 40-50K: No

| Income = 30-40K

| | Wine = Yes: No

| | Wine = No: Yes

| Income = 50-60K: No

| Income = 20-30K: No

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 14 70 %

Incorrectly Classified Instances 5 25 %

Kappa statistic 0.4509

Mean absolute error 0.2632

Root mean squared error 0.513

Relative absolute error 57.2082 %

Root relative squared error 106.8496 %

UnClassified Instances 1 5 %

Total Number of Instances 20

=== Detailed Accuracy By Class ===

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

0.714 0.250 0.625 0.714 0.667 0.454 0.688 0.541 Yes

0.750 0.286 0.818 0.750 0.783 0.454 0.750 0.764 No

Weighted Avg. 0.737 0.273 0.747 0.737 0.740 0.454 0.727 0.681

=== Confusion Matrix ===

a b <-- classified as

5 2 | a = Yes

3 9 | b = No

2)

Based on my decision tree I believe the company has conducted its promotions effectively, but they could definitely be improved. There is a 70% of correctly classified instances which is a low ratio. There is a 25% of incorrectly classified instances which means one every four instances will be classified incorrectly. This is a big percentage for a company wanting to maximise its profits, so we can see that it has not conducted its promotions as effectively as it could. As we can see from the tree the F-measure has a value of 0.740. This is because both the Precision and Recall have high values of 0.7+. As the F-measure can take values between 0 and 1, a value above 0.5 is considered good. Nevertheless as a company is supposed to have really accurate promotions in order to maximise profits, in my opinion it should have an F-measure between 0.8-0.9. As the Root Mean Squared Error (RMSE) is significantly higher than the Mean Absolute Error (MAE) we can assume that there are some outliers in our data which might interfere with our statistics.

3)

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and test set to evaluate it. This statistical method evaluates and compares learning algorithms by dividing the data into two sets (training and validating). In the 10-fold cross-validation, the original sample is randomly partitioned into 10 equal size subsamples. From those 10 subsamples one is used as the validation data for the testing model and the remaining 9 are used as training data. The cross-validation is then repeated 10 times (as many as the folds) and each of the subsamples is used once as the validation data. The 10 results are then averaged to produce a single estimation. This method allows us to have all observations as both training and validation data, while each observation is used for validation exactly one time, creating a less biased estimation.