

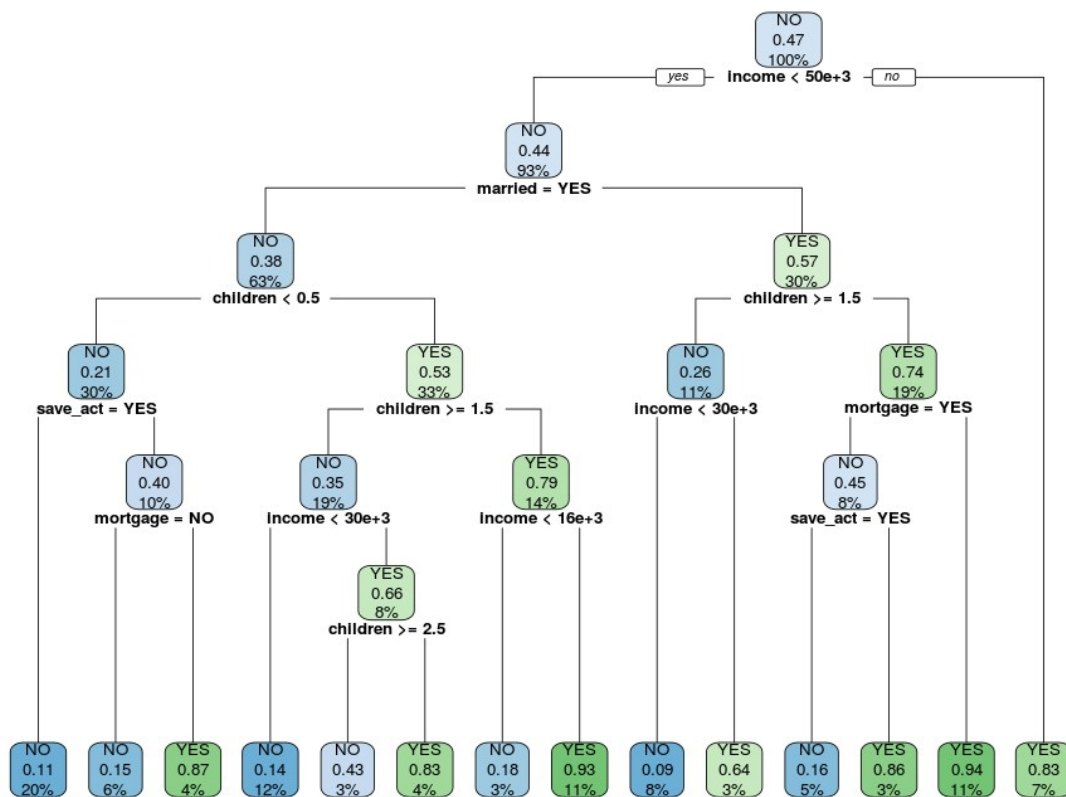
UW:450
Anish Mohan
04/30/2017

1. Model Accuracy:

Method	Test Set Accuracy	Test Set AUC
Decision Tree	90%	89.8%
Bagged Tree	91.67%	91.89 %
Random Forest	92.2%	92.08%

2. Used parameter sweeping and found the best complexity parameter for the tree – 0.005

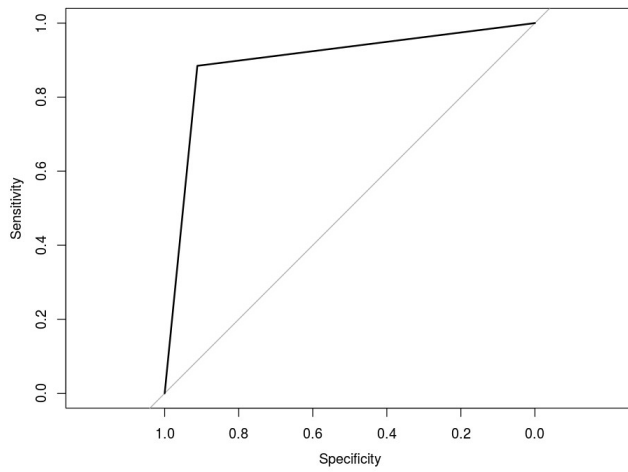
3. Tree:



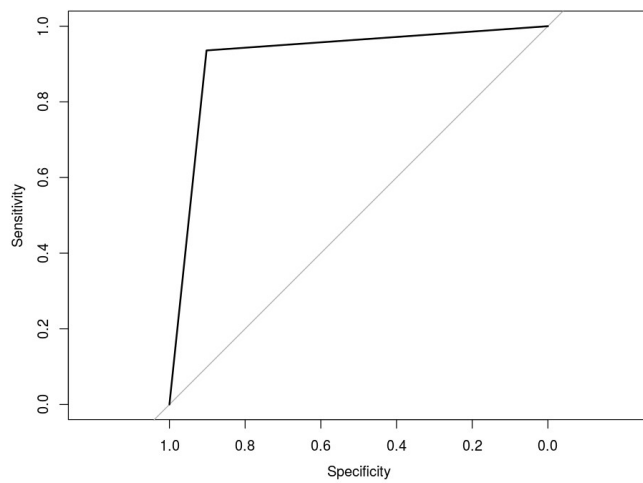
4. The Key parameters for the loan decisions were: Income, Children , Married and Mortgage.

5. ROC Curves shown for Decision Tree, Bagged Tree and Random Forest:

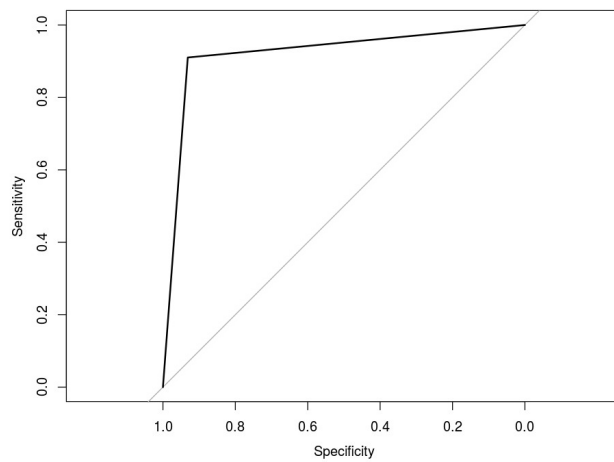
a) Decision Tree AUC



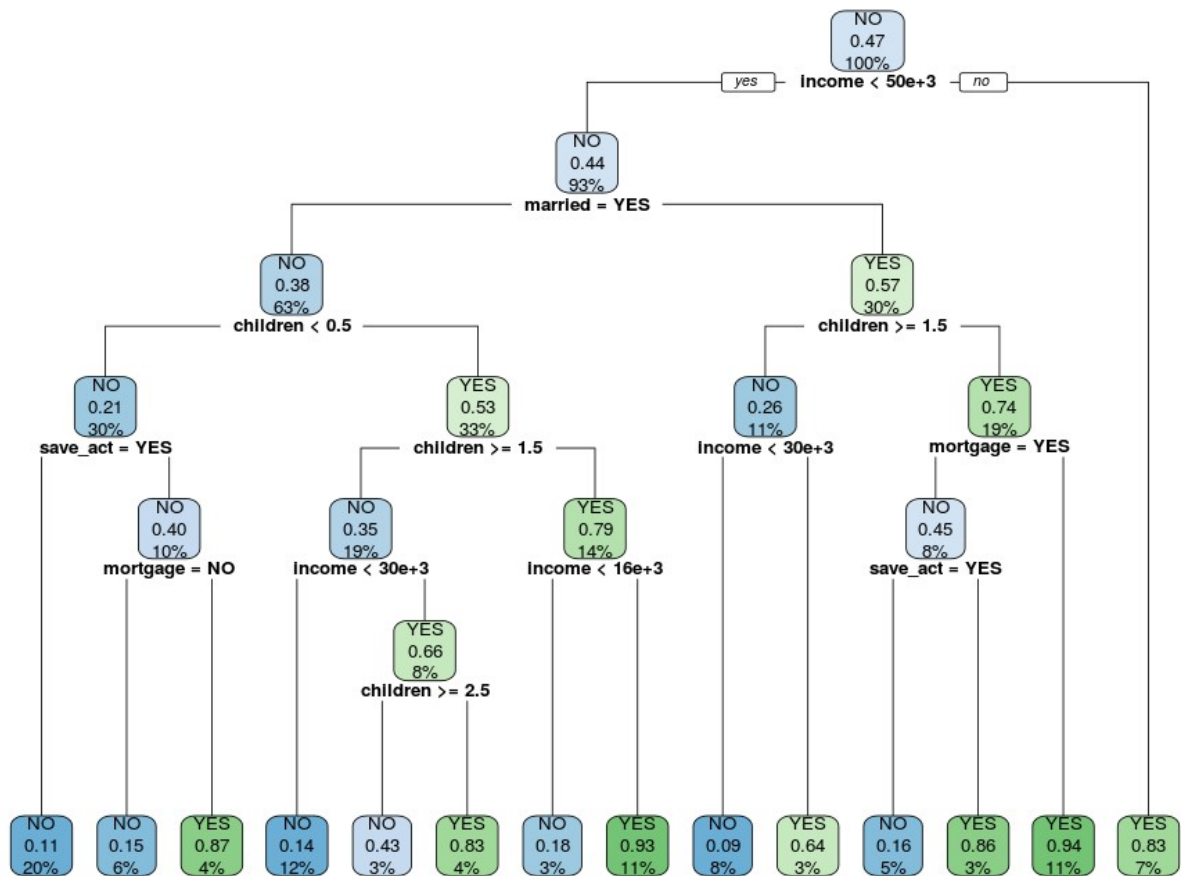
b) Bagged Tree ROC



c) Random Forest ROC



6)



Results:

Confusion Matrix and Statistics

```
predict.bankTest NO YES
                 NO  93   9
                 YES   9  69
```

```
Accuracy : 0.9
95% CI : (0.8466, 0.9396)
No Information Rate : 0.5667
P-Value [Acc > NIR] : <2e-16
```

```
Kappa : 0.7964
McNemar's Test P-Value : 1
```

```
Sensitivity : 0.9118
Specificity : 0.8846
Pos Pred Value : 0.9118
Neg Pred Value : 0.8846
Prevalence : 0.5667
Detection Rate : 0.5167
```

Detection Prevalence : 0.5667
Balanced Accuracy : 0.8982

'Positive' Class : NO

Full Results

Call:

```
rpart(formula = pep ~ ., data = bank.train, method = "class",  
      control = rpart.control(xval = 10, cp = 0.005))  
n= 420
```

	CP	nsplit	rel error	xerror	xstd
1	0.10204082	0	1.0000000	1.0000000	0.05216405
2	0.09948980	1	0.8979592	0.8775510	0.05141731
3	0.08418367	3	0.6989796	0.8520408	0.05117263
4	0.05102041	5	0.5306122	0.6224490	0.04746870
5	0.03571429	6	0.4795918	0.5663265	0.04610627
6	0.03316327	7	0.4438776	0.5204082	0.04483663
7	0.02806122	9	0.3775510	0.4846939	0.04374443
8	0.02040816	11	0.3214286	0.4081633	0.04105856
9	0.01020408	12	0.3010204	0.3469388	0.03851635
10	0.00500000	13	0.2908163	0.3571429	0.03896748

Variable importance

children	income	mortgage	age	save_act	married	region
31	27	14	11	10	4	3

Node number 1: 420 observations, complexity param=0.1020408
predicted class=NO expected loss=0.4666667 P(node) =1

class counts: 224 196

probabilities: 0.533 0.467

left son=2 (390 obs) right son=3 (30 obs)

Primary splits:

income	< 50163.6	to the left,	improve=8.687179, (0 missing)
married	splits as	RL,	improve=7.072920, (0 missing)
children	< 2.5	to the right,	improve=6.906667, (0 missing)
age	< 51.5	to the left,	improve=4.279142, (0 missing)
save_act	splits as	RL,	improve=2.159987, (0 missing)

Node number 2: 390 observations, complexity param=0.0994898

predicted class=NO expected loss=0.4384615 P(node) =0.9285714

class counts: 219 171

probabilities: 0.562 0.438

left son=4 (263 obs) right son=5 (127 obs)

Primary splits:

married	splits as	RL,	improve=6.216269, (0 missing)
children	< 2.5	to the right,	improve=6.157112, (0 missing)
save_act	splits as	RL,	improve=4.228063, (0 missing)
income	< 29669.75	to the left,	improve=3.239516, (0 missing)
age	< 30.5	to the left,	improve=2.441923, (0 missing)

Surrogate splits:

income	< 9632.49	to the right,	agree=0.687, adj=0.039, (0 split)
--------	-----------	---------------	-----------------------------------

Node number 3: 30 observations

predicted class=YES expected loss=0.1666667 P(node) =0.07142857

class counts: 5 25
probabilities: 0.167 0.833

Node number 4: 263 observations, complexity param=0.08418367
predicted class=NO expected loss=0.3764259 P(node) =0.6261905
class counts: 164 99
probabilities: 0.624 0.376
left son=8 (125 obs) right son=9 (138 obs)

Primary splits:

children < 0.5	to the left,	improve=13.515560, (0 missing)
income < 15052.9	to the left,	improve= 2.163421, (0 missing)
age < 30.5	to the left,	improve= 2.103210, (0 missing)
sex splits as LR,		improve= 1.463285, (0 missing)
mortgage splits as LR,		improve= 1.178720, (0 missing)

Surrogate splits:

income < 37556.3	to the right,	agree=0.555, adj=0.064, (0 split)
sex splits as LR,		agree=0.548, adj=0.048, (0 split)
region splits as RRLR,		agree=0.544, adj=0.040, (0 split)
save_act splits as LR,		agree=0.536, adj=0.024, (0 split)
age < 18.5	to the left,	agree=0.532, adj=0.016, (0 split)

Node number 5: 127 observations, complexity param=0.0994898
predicted class=YES expected loss=0.4330709 P(node) =0.302381
class counts: 55 72
probabilities: 0.433 0.567
left son=10 (46 obs) right son=11 (81 obs)

Primary splits:

children < 1.5	to the right,	improve=13.511960, (0 missing)
mortgage splits as RL,		improve= 4.390135, (0 missing)
save_act splits as RL,		improve= 2.920646, (0 missing)
income < 30158.7	to the left,	improve= 1.633877, (0 missing)
age < 27.5	to the left,	improve= 1.433257, (0 missing)

Node number 8: 125 observations, complexity param=0.02806122
predicted class=NO expected loss=0.208 P(node) =0.297619
class counts: 99 26
probabilities: 0.792 0.208
left son=16 (83 obs) right son=17 (42 obs)

Primary splits:

save_act splits as RL,		improve=4.897712, (0 missing)
mortgage splits as LR,		improve=4.601470, (0 missing)
age < 22.5	to the right,	improve=2.746998, (0 missing)
income < 30343.25	to the right,	improve=2.212571, (0 missing)
region splits as LLRL,		improve=1.023450, (0 missing)

Surrogate splits:

age < 18.5	to the right,	agree=0.672, adj=0.024, (0 split)
------------	---------------	-----------------------------------

Node number 9: 138 observations, complexity param=0.08418367
predicted class=YES expected loss=0.4710145 P(node) =0.3285714
class counts: 65 73
probabilities: 0.471 0.529
left son=18 (81 obs) right son=19 (57 obs)

Primary splits:

children < 1.5	to the right,	improve=13.1787700, (0 missing)
age < 30.5	to the left,	improve= 6.7675810, (0 missing)
income < 30332.9	to the left,	improve= 5.8986710, (0 missing)
car splits as RL,		improve= 0.5118134, (0 missing)
current_act splits as LR,		improve= 0.4946117, (0 missing)

Surrogate splits:

income < 8081.97 to the right, agree=0.601, adj=0.035, (0 split)
age < 19.5 to the right, agree=0.594, adj=0.018, (0 split)

Node number 10: 46 observations, complexity param=0.02040816
predicted class=NO expected loss=0.2608696 P(node) =0.1095238
class counts: 34 12
probabilities: 0.739 0.261
left son=20 (32 obs) right son=21 (14 obs)
Primary splits:
income < 30340.85 to the left, improve=5.8730590, (0 missing)
age < 28.5 to the left, improve=2.2097190, (0 missing)
region splits as LLRR, improve=1.2279740, (0 missing)
current_act splits as LR, improve=0.8352343, (0 missing)
children < 2.5 to the right, improve=0.6864989, (0 missing)
Surrogate splits:
age < 48.5 to the left, agree=0.783, adj=0.286, (0 split)

Node number 11: 81 observations, complexity param=0.03316327
predicted class=YES expected loss=0.2592593 P(node) =0.1928571
class counts: 21 60
probabilities: 0.259 0.741
left son=22 (33 obs) right son=23 (48 obs)
Primary splits:
mortgage splits as RL, improve=9.1224750, (0 missing)
save_act splits as RL, improve=3.7123530, (0 missing)
region splits as RRRL, improve=2.3625400, (0 missing)
income < 16999.25 to the left, improve=1.1201070, (0 missing)
age < 46.5 to the left, improve=0.7125397, (0 missing)
Surrogate splits:
age < 64.5 to the right, agree=0.617, adj=0.061, (0 split)

Node number 16: 83 observations
predicted class=NO expected loss=0.1084337 P(node) =0.197619
class counts: 74 9
probabilities: 0.892 0.108

Node number 17: 42 observations, complexity param=0.02806122
predicted class=NO expected loss=0.4047619 P(node) =0.1
class counts: 25 17
probabilities: 0.595 0.405
left son=34 (27 obs) right son=35 (15 obs)
Primary splits:
mortgage splits as LR, improve=9.9566140, (0 missing)
age < 53.5 to the left, improve=3.4380950, (0 missing)
sex splits as LR, improve=1.0252810, (0 missing)
region splits as LLRL, improve=0.9586835, (0 missing)
income < 21559.2 to the right, improve=0.8875070, (0 missing)
Surrogate splits:
age < 53.5 to the left, agree=0.714, adj=0.2, (0 split)

Node number 18: 81 observations, complexity param=0.05102041
predicted class=NO expected loss=0.345679 P(node) =0.1928571
class counts: 53 28
probabilities: 0.654 0.346
left son=36 (49 obs) right son=37 (32 obs)
Primary splits:
income < 30332.9 to the left, improve=10.204480, (0 missing)
region splits as LRRL, improve= 2.620045, (0 missing)
mortgage splits as RL, improve= 1.895349, (0 missing)

age < 46.5 to the left, improve= 1.224393, (0 missing)
children < 2.5 to the right, improve= 1.202760, (0 missing)
Surrogate splits:
age < 42 to the left, agree=0.753, adj=0.375, (0 split)
region splits as LRRL, agree=0.630, adj=0.062, (0 split)

Node number 19: 57 observations, complexity param=0.03571429
predicted class=YES expected loss=0.2105263 P(node) =0.1357143
class counts: 12 45
probabilities: 0.211 0.789
left son=38 (11 obs) right son=39 (46 obs)
Primary splits:
income < 15576.45 to the left, improve=10.0659500, (0 missing)
age < 30 to the left, improve= 6.9374020, (0 missing)
car splits as RL, improve= 1.3537720, (0 missing)
region splits as LRLR, improve= 1.1775470, (0 missing)
save_act splits as LR, improve= 0.7935223, (0 missing)
Surrogate splits:
age < 22.5 to the left, agree=0.877, adj=0.364, (0 split)

Node number 20: 32 observations
predicted class=NO expected loss=0.09375 P(node) =0.07619048
class counts: 29 3
probabilities: 0.906 0.094

Node number 21: 14 observations
predicted class=YES expected loss=0.3571429 P(node) =0.03333333
class counts: 5 9
probabilities: 0.357 0.643

Node number 22: 33 observations, complexity param=0.03316327
predicted class=NO expected loss=0.4545455 P(node) =0.07857143
class counts: 18 15
probabilities: 0.545 0.455
left son=44 (19 obs) right son=45 (14 obs)
Primary splits:
save_act splits as RL, improve=7.8824330, (0 missing)
region splits as RLRL, improve=1.9414140, (0 missing)
income < 16531.75 to the left, improve=1.7262740, (0 missing)
sex splits as RL, improve=0.9251748, (0 missing)
age < 40.5 to the right, improve=0.8080808, (0 missing)
Surrogate splits:
age < 35.5 to the right, agree=0.758, adj=0.429, (0 split)
region splits as RLRL, agree=0.697, adj=0.286, (0 split)
income < 26809.8 to the right, agree=0.606, adj=0.071, (0 split)

Node number 23: 48 observations
predicted class=YES expected loss=0.0625 P(node) =0.1142857
class counts: 3 45
probabilities: 0.062 0.938

Node number 34: 27 observations
predicted class=NO expected loss=0.1481481 P(node) =0.06428571
class counts: 23 4
probabilities: 0.852 0.148

Node number 35: 15 observations
predicted class=YES expected loss=0.1333333 P(node) =0.03571429
class counts: 2 13

probabilities: 0.133 0.867

Node number 36: 49 observations
predicted class=NO expected loss=0.1428571 P(node) =0.1166667
class counts: 42 7
probabilities: 0.857 0.143

Node number 37: 32 observations, complexity param=0.01020408
predicted class=YES expected loss=0.34375 P(node) =0.07619048
class counts: 11 21
probabilities: 0.344 0.656
left son=74 (14 obs) right son=75 (18 obs)

Primary splits:
children < 2.5 to the right, improve=2.5803570, (0 missing)
income < 44172.05 to the left, improve=2.1175000, (0 missing)
mortgage splits as RL, improve=1.1234900, (0 missing)
age < 57.5 to the right, improve=0.6075405, (0 missing)
region splits as LRRR, improve=0.6075405, (0 missing)
Surrogate splits:
region splits as RLRR, agree=0.719, adj=0.357, (0 split)
income < 48845.95 to the right, agree=0.656, adj=0.214, (0 split)
age < 56 to the right, agree=0.625, adj=0.143, (0 split)
save_act splits as RL, agree=0.625, adj=0.143, (0 split)
current_act splits as LR, agree=0.594, adj=0.071, (0 split)

Node number 38: 11 observations
predicted class=NO expected loss=0.1818182 P(node) =0.02619048
class counts: 9 2
probabilities: 0.818 0.182

Node number 39: 46 observations
predicted class=YES expected loss=0.06521739 P(node) =0.1095238
class counts: 3 43
probabilities: 0.065 0.935

Node number 44: 19 observations
predicted class=NO expected loss=0.1578947 P(node) =0.0452381
class counts: 16 3
probabilities: 0.842 0.158

Node number 45: 14 observations
predicted class=YES expected loss=0.1428571 P(node) =0.03333333
class counts: 2 12
probabilities: 0.143 0.857

Node number 74: 14 observations
predicted class=NO expected loss=0.4285714 P(node) =0.03333333
class counts: 8 6
probabilities: 0.571 0.429

Node number 75: 18 observations
predicted class=YES expected loss=0.1666667 P(node) =0.04285714
class counts: 3 15
probabilities: 0.167 0.833