In [1]:
```
using StatsBase, Random, LinearAlgebra, Plots, DecisionTree, DataFrames, ML
```

In [2]:
```
data = CSV.read("stevens.csv", DataFrame, header=1, pool=true);
```

In [4]:
```
y = data[:, 7]
X = data[:, 1:6]
(train_X, train_y), (test_X, test_y) = IAI.split_data(:classification, X, y
```

a)

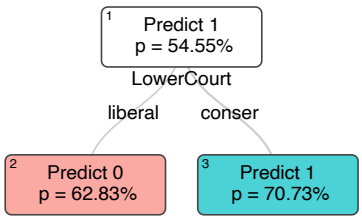
All models had identical training accuracy and AUC, with the exception of the entropy based model having slightly improved AUC. Additionally, all models had effectively the same out-of-sample accuracy. Again, the entropy based model had slightly higher AUC as opposed to the other models. These nearly identical when rounded performance metrics are a rarity and we would not expect this to happen often. This is a result of the misclassification and gini measures creating identical trees.

In [5]:
```
lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:misclassifica
grid = IAI.GridSearch(lnr)
IAI.fit!(grid, train_X, train_y, validation_criterion= :auc)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:00
4m  Parameters:  cp=>0.08069
```

Out[5]:

Collapse Expand Save PNG



In [6]:
```
println("In-Sample Accuracy = ", round(IAI.score(lnr,train_X, train_y,crite
println("In-Sample AUC = ", round(IAI.score(lnr,train_X, train_y,criterion=
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr,test_X, test_y,cri
println("Out-of-Sample AUC = ", round(IAI.score(lnr,test_X, test_y,criterio
```

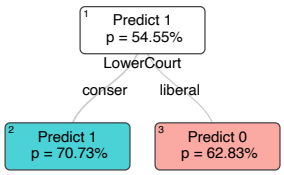
```
In-Sample Accuracy = 0.6692
In-Sample AUC = 0.669
Out-of-Sample Accuracy = 0.6647
Out-of-Sample AUC = 0.6645
```

In [7]:
```
lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:gini, max_dep
grid = IAI.GridSearch(lnr)
IAI.fit!(grid, train_X, train_y, validation_criterion= :auc)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:00
4m  Parameters:  cp=>0.06777
```

Out[7]:

Collapse Expand Save PNG



In [8]:
```
println("In-Sample Accuracy = ", round(IAI.score(lnr,train_X, train_y,crite
println("In-Sample AUC = ", round(IAI.score(lnr,train_X, train_y,criterion=
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr,test_X, test_y,cri
println("Out-of-Sample AUC = ", round(IAI.score(lnr,test_X, test_y,criterio
```

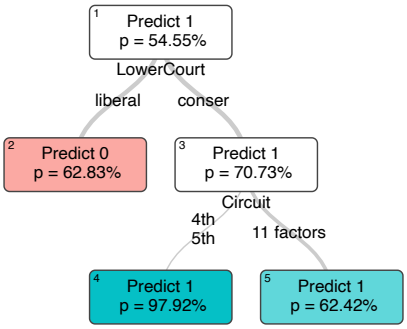
```
In-Sample Accuracy = 0.6692
In-Sample AUC = 0.669
Out-of-Sample Accuracy = 0.6647
Out-of-Sample AUC = 0.6645
```

In [9]:
```
lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:entropy, max_
grid = IAI.GridSearch(lnr)
IAI.fit!(grid, train_X, train_y, validation_criterion= :auc)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████████| Time: 0:00:00
4m  Parameters:  cp=>0.03886
```

Out[9]:

Collapse Expand Save PNG



In [10]:
```
println("In-Sample Accuracy = ", round(IAI.score(lnr, train_X, train_y, cri
println("In-Sample AUC = ", round(IAI.score(lnr, train_X, train_y, criterio
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
In-Sample Accuracy = 0.6692
In-Sample AUC = 0.7034
Out-of-Sample Accuracy = 0.6647
Out-of-Sample AUC = 0.6676
```

b)

Neither accuracy nor AUC converged on a stable optimal value or provided smooth functions. However, both metrics had local optimal values at a depth of twenty-five. Based on only that

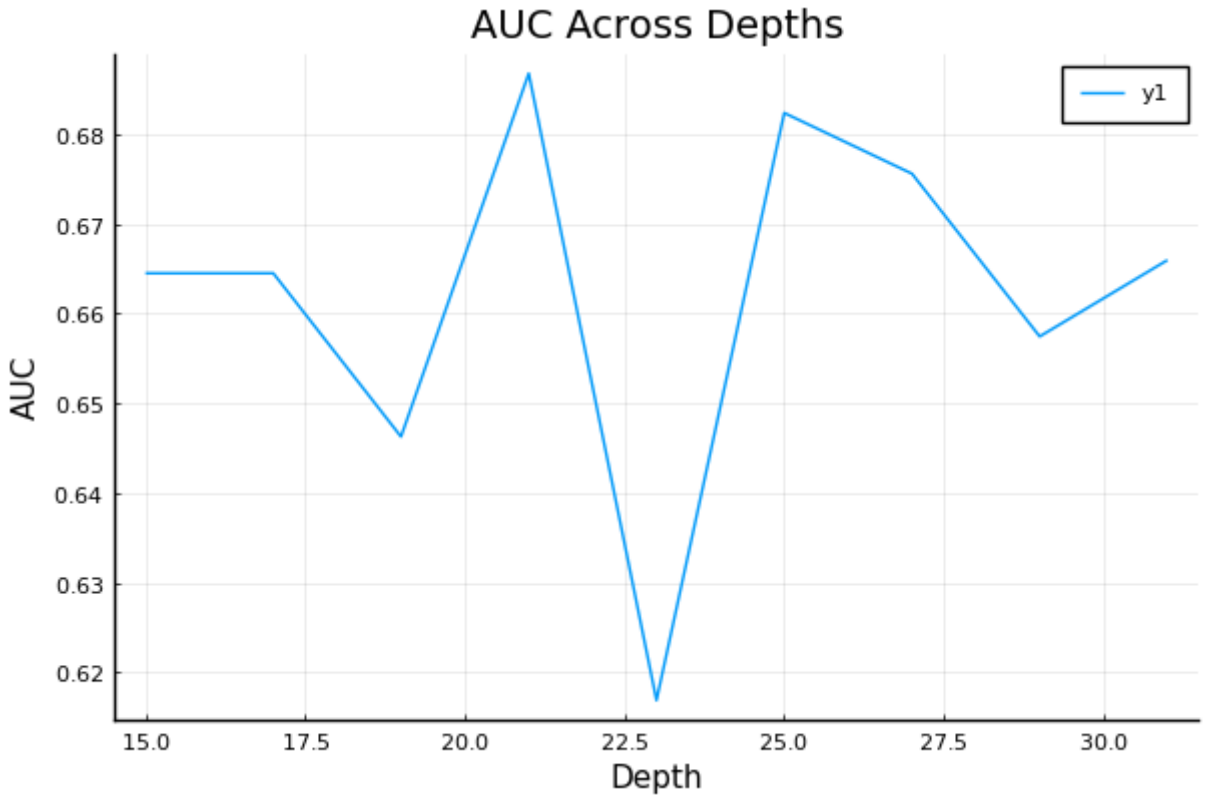
information, I would likely select a depth of twenty-five for the tree. On the other hand, the complexity parameter converged to a very small value once it reached a depth of 19.

In [62]:
```
auc = []
accuracy = []
cp = []
dep =[15,17,19,21,23,25,27,29,31]

for i=1:length(dep)
    lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:gini, max
    grid = IAI.GridSearch(lnr)
    IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
    lnr = IAI.get_learner(grid)

    append!(auc,IAI.score(lnr,test_X, test_y,criterion=:auc))
    append!(accuracy,IAI.score(lnr,test_X, test_y,criterion=:misclassificat
    append!(cp,lnr.cp)
end
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:02
4m  Parameters:  cp=>0.0009074
```

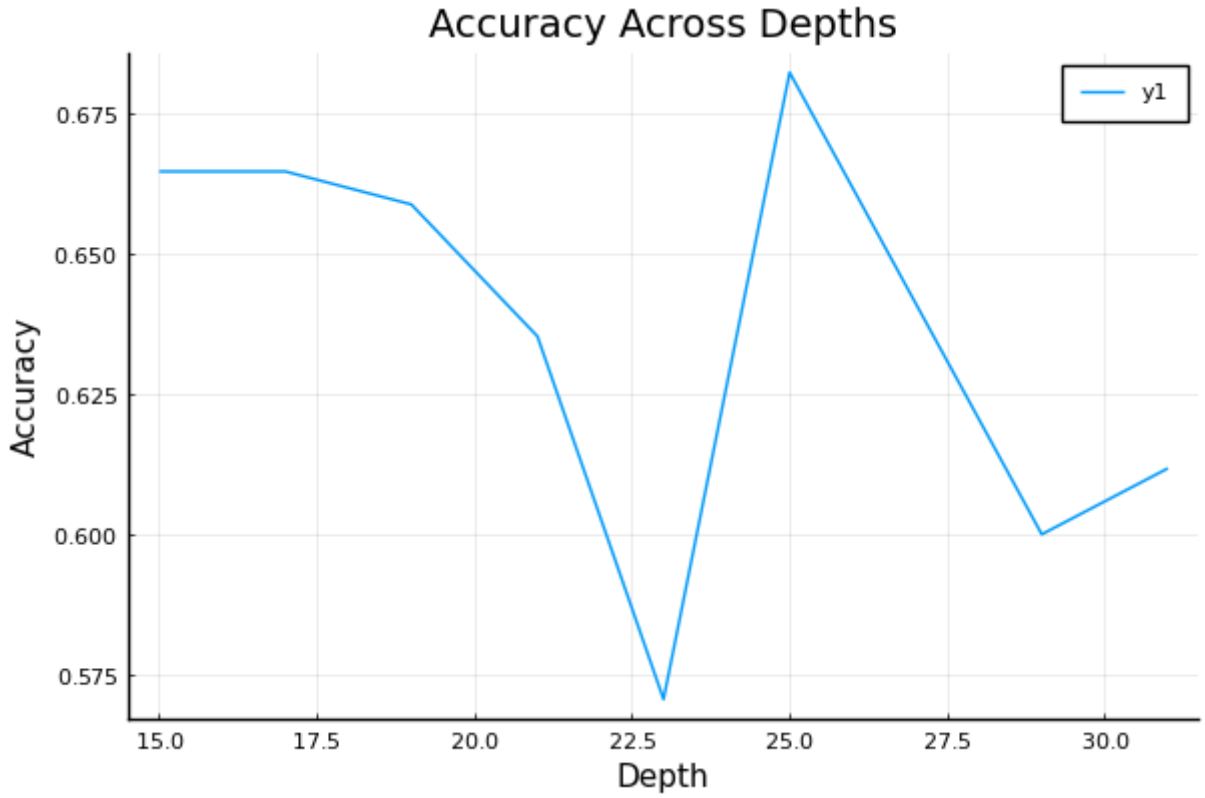
In [63]:
```
pyplot()
plot(dep, auc, title = "AUC Across Depths", ylabel = "AUC", xlabel = "Depth
```

Out[63]:



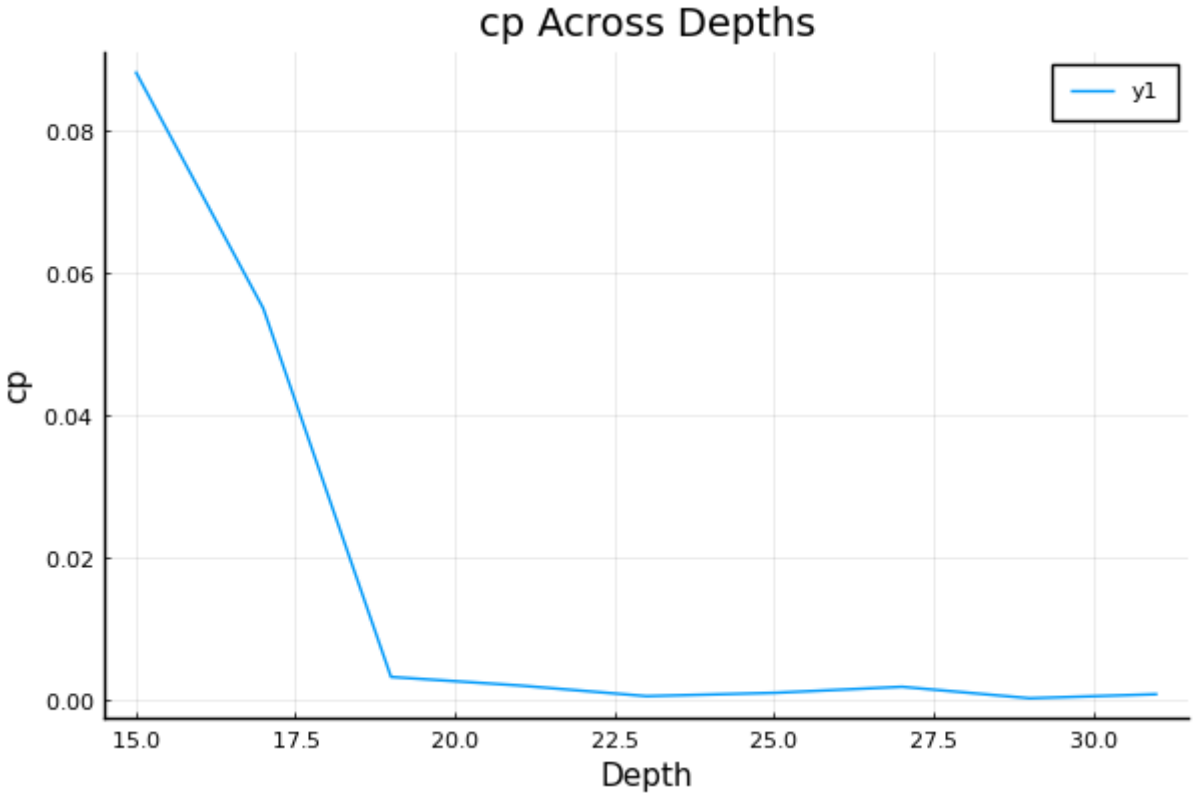
In [64]:
```
pyplot()
plot(dep, accuracy, title = "Accuracy Across Depths", ylabel = "Accuracy",
```

Out[64]:



In [65]:
```
plot(dep, cp, title = "cp Across Depths", ylabel = "cp", xlabel = "Depth")
```

Out[65]:



c)

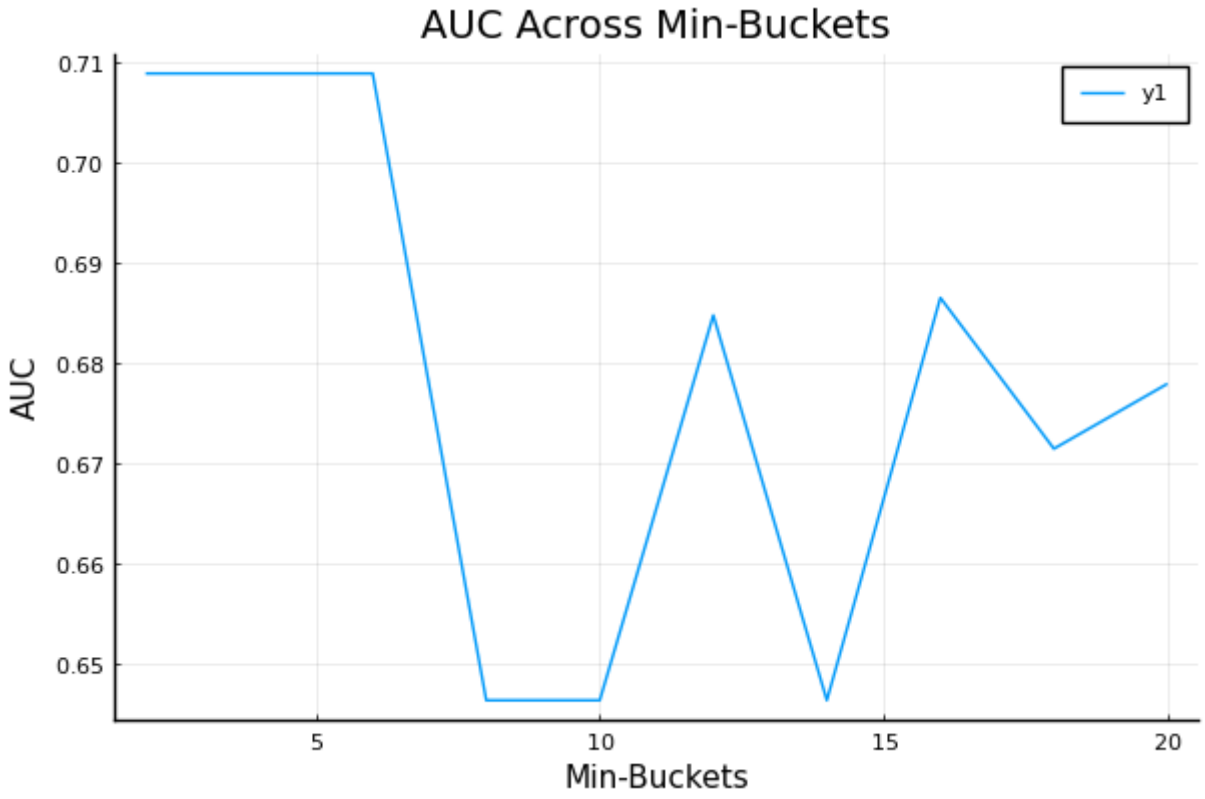
Both out-of-sample accuracy and AUC have their highest values with the lowest threshold for minimum number of observations per bucket. In an effort to prevent overfitting, I would select the highest value of min-bucket that still optimizes accuracy and AUC. In this case, I would select 6 for min-buckets. In this case, the complexity parameter converged to a very small value for low values of min-bucket.

In [15]:
```
auc = []
accuracy = []
cp = []
min_buck =[2, 4, 6, 8, 10, 12, 14, 16, 18, 20]

for i=1:length(dep)
    lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:gini, max
    grid = IAI.GridSearch(lnr)
    IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
    lnr = IAI.get_learner(grid)

    append!(auc,IAI.score(lnr,test_X, test_y,criterion=:auc))
    append!(accuracy,IAI.score(lnr,test_X, test_y,criterion=:misclassificat
    append!(cp,lnr.cp)
end
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│  - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│  - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:00
4m  Parameters:  cp=>0.006245
```

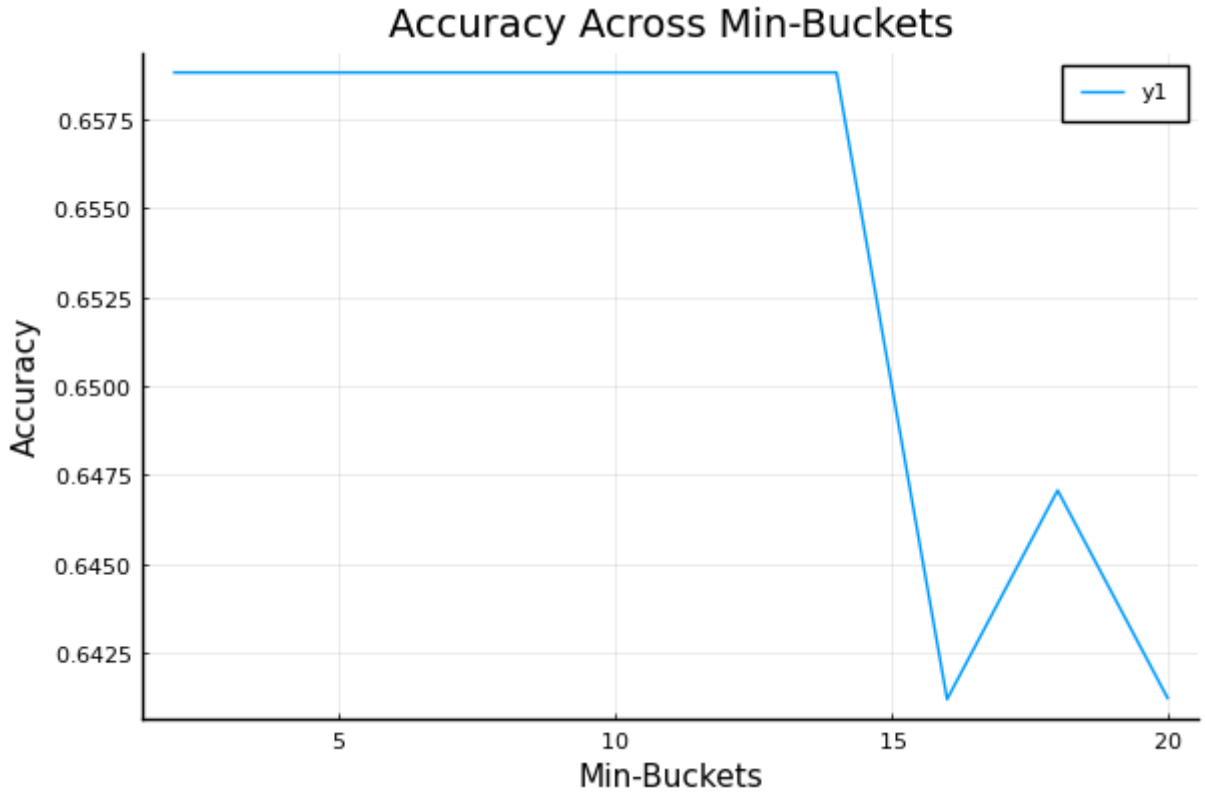
In [16]:
```
pyplot()
plot(min_buck, auc, title = "AUC Across Min-Buckets", ylabel = "AUC", xlabe
```

Out[16]:



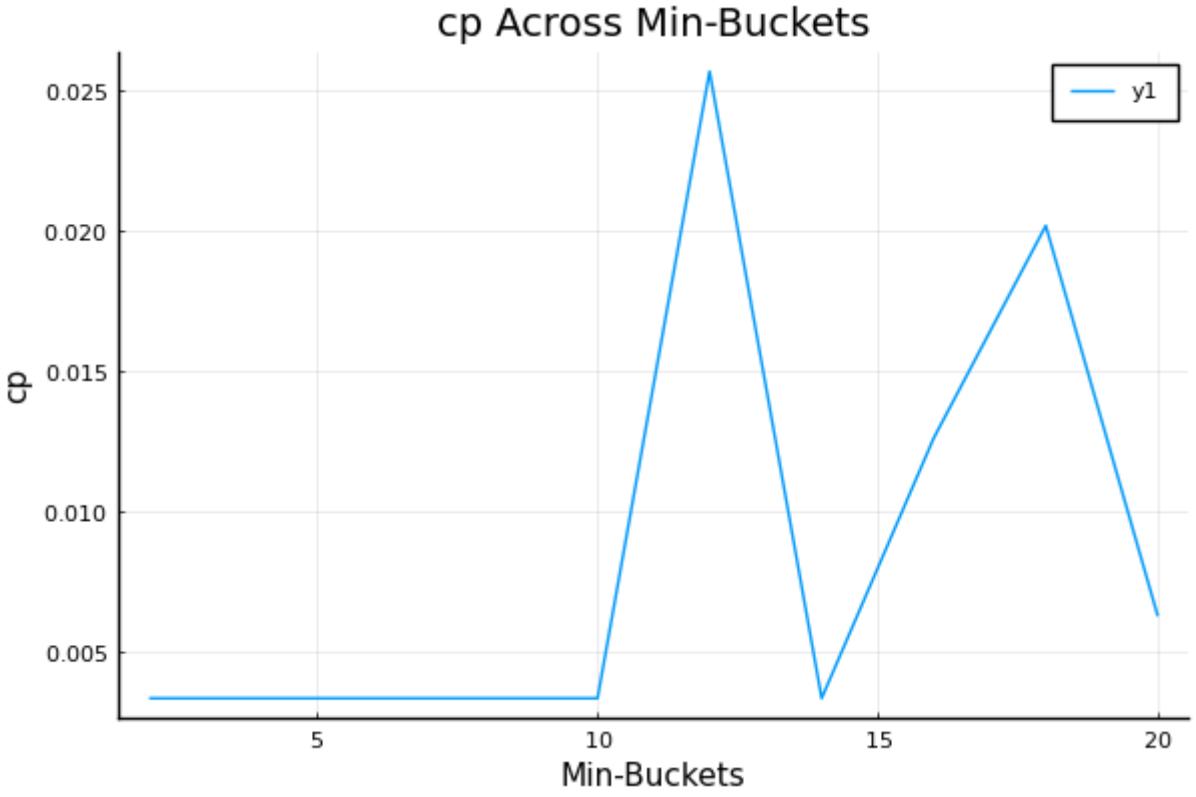
In [17]:
```
pyplot()
plot(min_buck, accuracy, title = "Accuracy Across Min-Buckets", ylabel = "A
```

Out[17]:



In [18]:
```
plot(min_buck, cp, title = "cp Across Min-Buckets", ylabel = "cp", xlabel =
```

Out[18]:



d)

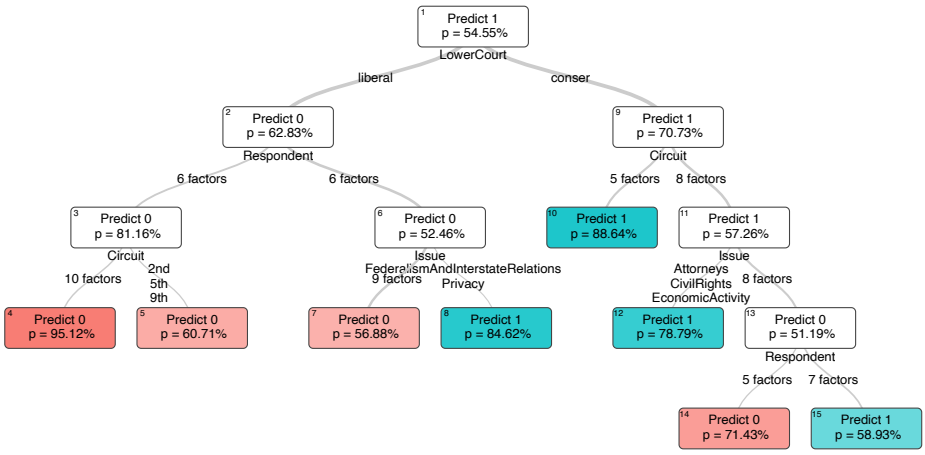
CART

In [19]:
```
lnr = IAI.OptimalTreeClassifier(random_seed=15095, localsearch=false, crite
grid = IAI.GridSearch(lnr, max_depth=[2,4], minbucket=5:10)
IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████████| Time: 0:00:00
4m  Parameters:  minbucket=>7 cp=>0.01398 max_depth=>4
```

Out[19]:

Collapse Expand Save PNG



In [20]:
```
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
Out-of-Sample Accuracy = 0.5941
Out-of-Sample AUC = 0.6115
```

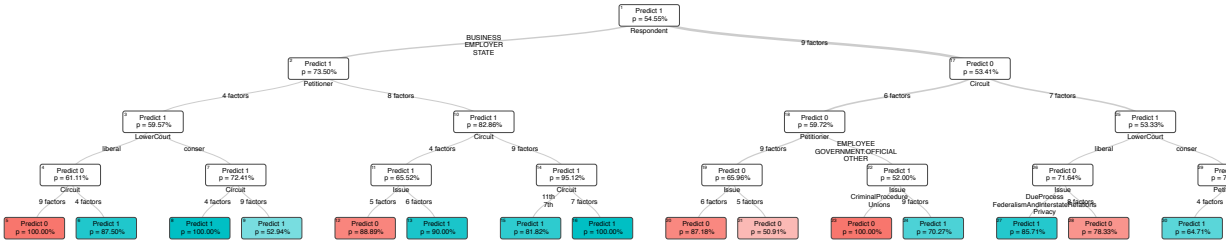
OCT

In [27]:
```
lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:gini)
grid = IAI.GridSearch(lnr, max_depth=[2,4], minbucket=5:10)
IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████████| Time: 0:00:00
4m  Parameters:  minbucket=>7 cp=>0.001371 max_depth=>4
```

Out[27]:

Collapse Expand Save PNG



In [28]:
```
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
Out-of-Sample Accuracy = 0.6412
Out-of-Sample AUC = 0.6533
```

Random Forest

In [23]:
```
lnr = IAI.RandomForestClassifier(random_seed=15095, criterion=:gini)
grid = IAI.GridSearch(lnr, max_depth=[2,4], minbucket=5:10, num_trees=[50,1
IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:00
4m  Parameters:  minbucket=>8 num_trees=>50 max_depth=>4
```

Out[23]:
```
Fitted RandomForestClassifier
```

In [24]:
```
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
Out-of-Sample Accuracy = 0.6882
Out-of-Sample AUC = 0.735
```

Boosted Tree

In [31]:
```
lnr = IAI.XGBoostClassifier(random_seed=15095, criterion=:entropy)
grid = IAI.GridSearch(lnr, max_depth=[2,4,6], minbucket=5:10, num_estimator
IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:00
4m  Parameters:  minbucket=>5 num_estimators=>50 max_depth=>2
```

Out[31]:
```
Fitted XGBoostClassifier
```

In [33]:
```
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
Out-of-Sample Accuracy = 0.6412
Out-of-Sample AUC = 0.682
```

Sparse Logistic Regression

In [53]:
```
lnr = IAI.OptimalFeatureSelectionClassifier(random_seed=15095, criterion=:e
grid = IAI.GridSearch(lnr, sparsity=1:10,gamma=[.001,0.01,0.1,0.25,.5,1,2,5
IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│  - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│  - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████| Time: 0:00:00
4m  Parameters:  gamma=>2 sparsity=>1
```

Out[53]:
```
Fitted OptimalFeatureSelectionClassifier:
  Constant: 0.882576
  Weights:
    LowerCourt==liberal: -1.38754
  (Higher score indicates stronger prediction for class `1`)
```

In [54]:
```
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
Out-of-Sample Accuracy = 0.6647
Out-of-Sample AUC = 0.6645
```

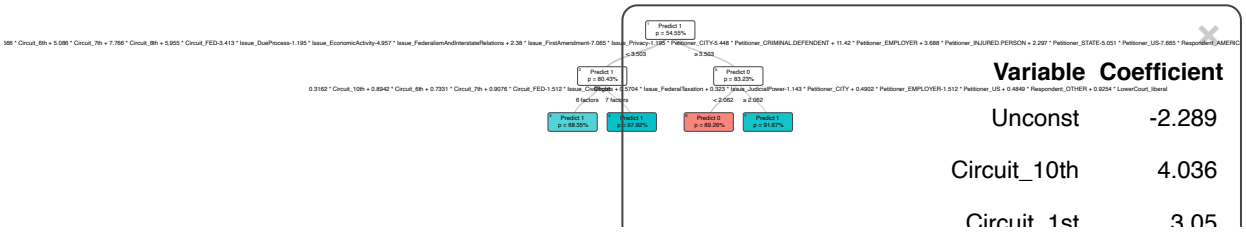
Additional Model: OCT-H

In [55]:
```
lnr = IAI.OptimalTreeClassifier(random_seed=15095, criterion=:gini, hyperpl
grid = IAI.IAI.GridSearch(lnr, max_depth=1:2, cp = cp = [0.001, 0.005] )
IAI.fit_cv!(grid, train_X, train_y, n_folds=5)
lnr = IAI.get_learner(grid)
```

```
┌ Warning: ProgressMeter by default refresh meters with additional inform
ation in IJulia via `IJulia.clear_output`, which clears all outputs in th
e cell.
│   - To prevent this behaviour, do `ProgressMeter.ijulia_behavior(:appen
d)`.
│   - To disable this warning message, do `ProgressMeter.ijulia_behavior(:
clear)`.
└ @ ProgressMeter /Users/iai/builds/InterpretableAI/SystemImage/SysImgBui
lder/.julia/packages/ProgressMeter/Vf8un/src/ProgressMeter.jl:620
Refitting with best parameters...  100%|████████████████████| Time: 0:00:29
4m  Parameters:  cp=>0.005 max_depth=>2
```

Out[55]:

Collapse Expand Save PNG



In [57]:
```
println("Out-of-Sample Accuracy = ", round(IAI.score(lnr, test_X, test_y, c
println("Out-of-Sample AUC = ", round(IAI.score(lnr, test_X, test_y, criter
```

```
Out-of-Sample Accuracy = 0.6706
Out-of-Sample AUC = 0.6593
```