Specific Kaggle Submission Code is towards the bottom

Trial and Error: RPart

```
> tree <- rpart(Hired ~ Major+TwitterFOLLOWERS, data = HireRTrain1.1,method =
"class", control = rpart.control(minbucket = 200))

> rpart.plot(tree)

> tree <- rpart(Hired ~ Major+TwitterFOLLOWING, data = HireRTrain1.1,method =
"class", control = rpart.control(minbucket = 200))

> rpart.plot(tree)

> tree <- rpart(Hired ~ Major+TikTokFOLLOWERS, data = HireRTrain1.1,method =
"class", control = rpart.control(minbucket = 200))

> rpart.plot(tree)

> tree <- rpart(Hired ~ Major+TikTokTFOLLOWING, data = HireRTrain1.1,method =
"class", control = rpart.control(minbucket = 200))

> rpart.plot(tree)

> tree <- rpart(Hired ~ TwitterFOLLOWERS+Coding, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)


> tree <- rpart(Hired ~ TwitterFOLLOWING+Coding, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)

> tree <- rpart(Hired ~ TikTokFOLLOWERS+Coding, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)

> tree <- rpart(Hired ~ TikTokTFOLLOWING+Coding, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)

> tree <- rpart(Hired ~ TwitterFOLLOWERS+TwitterFOLLOWING,
data = HireRTrain1.1,method = "class")

> rpart.plot(tree)
```

```
> tree <- rpart(Hired ~ Impression+Major+TikTokTFOLLOWING, data =
HireRTrain1.1,method = "class")

> rpart.plot(tree)

> tree <- rpart(Hired ~ Impression+Major+TwitterFOLLOWERS, data =
HireRTrain1.1,method = "class")
> rpart.plot(tree)



> tree <- rpart(Hired ~ College+Impression, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)

> tree <- rpart(Hired ~ College+TwitterFOLLOWERS, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)


> tree <- rpart(Hired ~ College+TwitterFOLLOWING, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)
> tree <- rpart(Hired ~ College+TikTokTFOLLOWING, data = HireRTrain1.1,method =
"class")

> rpart.plot(tree)
```

Trial and Error: Cross Validation

```
> tree <- rpart(Hired ~
Coding+Impression++TikTokFOLLOWERS+TwitterFOLLOWING+TikTokTFOLLOWING+TwitterFOL
LOWERS, data = HireRTrain1.1,method = "class")
> pred <- predict(tree, HireRTrain1.1, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No  No  No Yes
Levels: No Yes
> mean(HireRTrain1.1$Hired == pred)
[1] 0.8738278
> CrossValidation::cross_validate(HireRTrain1.1, tree, 2, 0.7)
[[1]]
  accuracy_subset accuracy_all
1       0.8437500    0.8437500
2       0.8210227    0.8210227
```

```
[[2]]
[[2]]$average_accuracy_subset
[1] 0.8323864

[[2]]$average_accuracy_all
[1] 0.8323864

[[2]]$variance_accuracy_subset
[1] 0.0002582645

[[2]]$variance_accuracy_all
[1] 0.0002582645
```

```
> tree <- rpart(Hired ~
Coding+Impression+TwitterFOLLOWERS+TikTokFOLLOWERS+TwitterFOLLOWING+TikTokTFOLL
OWING, data = HireRTrain1.1,method = "class")
> pred <- predict(tree, HireRTrain1.1, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No  No  No Yes
Levels: No Yes
> mean(HireRTrain1.1$Hired == pred)
[1] 0.8738278
> CrossValidation::cross_validate(HireRTrain1.1, tree, 2, 0.7)
[[1]]
  accuracy_subset accuracy_all
1       0.8295455    0.8295455
2       0.8323864    0.8323864

[[2]]
[[2]]$average_accuracy_subset
[1] 0.8309659

[[2]]$average_accuracy_all
[1] 0.8309659

[[2]]$variance_accuracy_subset
[1] 4.035382e-06

[[2]]$variance_accuracy_all
[1] 4.035382e-06
```

```
> tree <- rpart(Hired ~
Coding+TikTokTFOLLOWING+Impression+TwitterFOLLOWERS+TikTokFOLLOWERS+TwitterFOL
LOWING+Major, data = HireRTrain1.1,method = "class")
> pred <- predict(tree, HireRTrain1.1, type = "class")
> head(pred)
```

```
  1   2   3   4   5   6
Yes Yes  No  No  No Yes
Levels: No Yes
> mean(HireRTrain1.1$Hired == pred)
[1] 0.8738278
> CrossValidation::cross_validate(HireRTrain1.1, tree, 2, 0.7)
[[1]]
  accuracy_subset accuracy_all
1       0.8494318    0.8494318
2       0.8380682    0.8380682

[[2]]
[[2]]$average_accuracy_subset
[1] 0.84375

[[2]]$average_accuracy_all
[1] 0.84375

[[2]]$variance_accuracy_subset
[1] 6.456612e-05

[[2]]$variance_accuracy_all
[1] 6.456612e-05
```

```
> tree <- rpart(Hired ~
Coding+TikTokTFOLLOWING+Impression+TwitterFOLLOWERS+TikTokFOLLOWERS, data =
HireRTrain1.1,method = "class")
> pred <- predict(tree, HireRTrain1.1, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No  No  No Yes
Levels: No Yes
> mean(HireRTrain1.1$Hired == pred)
[1] 0.8644501
> CrossValidation::cross_validate(HireRTrain1.1, tree, 2, 0.7)
[[1]]
  accuracy_subset accuracy_all
1       0.8295455    0.8295455
2       0.8551136    0.8806818

[[2]]
[[2]]$average_accuracy_subset
[1] 0.8423295

[[2]]$average_accuracy_all
[1] 0.8551136

[[2]]$variance_accuracy_subset
[1] 0.000326866
```

```
[[2]]$variance_accuracy_all
[1] 0.001307464
```

```
> tree <- rpart(Hired ~
Coding+Impression+TwitterFOLLOWING+TwitterFOLLOWERS+TikTokFOLLOWERS, data =
HireRTrain1.1,method = "class")
> pred <- predict(tree, HireRTrain1.1, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No  No  No Yes
Levels: No Yes
> mean(HireRTrain1.1$Hired == pred)
[1] 0.8635976
> CrossValidation::cross_validate(HireRTrain1.1, tree, 2, 0.7)
[[1]]
  accuracy_subset accuracy_all
1       0.8693182    0.8664773
2       0.8352273    0.8465909

[[2]]
[[2]]$average_accuracy_subset
[1] 0.8522727

[[2]]$average_accuracy_all
[1] 0.8565341

[[2]]$variance_accuracy_subset
[1] 0.000581095

[[2]]$variance_accuracy_all
[1] 0.0001977337
```

```
> tree <- rpart(Hired ~
Coding+TikTokTFOLLOWING+Impression+TwitterFOLLOWING+TwitterFOLLOWERS+TikTokFOLL
OWERS, data = HireRTrain1.1,method = "class")
> pred <- predict(tree, HireRTrain1.1, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No  No  No Yes
Levels: No Yes
> mean(HireRTrain1.1$Hired == pred)
[1] 0.8738278
> CrossValidation::cross_validate(HireRTrain1.1, tree, 2, 0.7)
[[1]]
  accuracy_subset accuracy_all
1       0.8437500    0.8437500
2       0.8380682    0.8380682
```

```
[[2]]
[[2]]$average_accuracy_subset
[1] 0.8409091

[[2]]$average_accuracy_all
[1] 0.8409091

[[2]]$variance_accuracy_subset
[1] 1.614153e-05

[[2]]$variance_accuracy_all
[1] 1.614153e-05
```

Kaggle Submission 1

```
> test_challenge2 <- read.csv("C:/Users/msraa/Downloads/test_challenge2.csv",
stringsAsFactors=TRUE)
>  View(test_challenge2)
> sample_submission2 <-
read.csv("C:/Users/msraa/Downloads/sample_submission2.csv",
stringsAsFactors=TRUE)
>  View(sample_submission2)
> submission <- sample_submission2

> library(rpart)
> library(rpart.plot)

> HireRTrain1.1 <- read.csv("C:/Users/msraa/Downloads/HireRTrain1-1.csv",
stringsAsFactors=TRUE)
>  View(HireRTrain1.1)

> tree <- rpart(Hired ~
Coding+TikTokTFOLLOWING+Impression+TwitterFOLLOWERS+TikTokFOLLOWERS+TwitterFOLL
OWING+Major, data = HireRTrain1.1,method = "class")
> rpart.plot(tree)
> pred <- predict(tree, test_challenge2, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No Yes Yes Yes
Levels: No Yes
> submission$Prediction <- pred
> nrow(submission)
[1] 779
> submission
[ reached 'max' / getOption("max.print") -- omitted 279 rows ]
> write.csv(submission, 'predictionChallenge2.csv', row.names = FALSE)
```

Kaggle Submission 2

```
> test_challenge2 <- read.csv("C:/Users/msraa/Downloads/test_challenge2.csv",
stringsAsFactors=TRUE)
>  View(test_challenge2)
> sample_submission2 <-
read.csv("C:/Users/msraa/Downloads/sample_submission2.csv",
stringsAsFactors=TRUE)
>  View(sample_submission2)
> HireTrainApr10 <- read.csv("C:/Users/msraa/Downloads/HireTrainApr10.csv",
stringsAsFactors=TRUE)
>  View(HireTrainApr10)

> library(rpart)
> library(rpart.plot)

> tree <- rpart(Hired ~
Coding+Impression++TikTokFOLLOWERS+TwitterFOLLOWING+TikTokTFOLLOWING+TwitterFOL
LOWERS, data = HireRTrain1.1,method = "class")
> rpart.plot(tree)
> pred <- predict(tree, test_challenge2, type = "class")
> head(pred)
  1   2   3   4   5   6
Yes Yes  No Yes Yes Yes
Levels: No Yes
> submission$Prediction <- pred
> nrow(submission)
[1] 779
> submission
```