R Notebook: random forest on dataset with classes - full modeal

```
library(caret)
## Loading required package: ggplot2
## Loading required package: lattice
library(ggplot2)
library(randomForest)
## randomForest 4.6-14
## Type rfNews() to see new features/changes/bug fixes.
## Attaching package: 'randomForest'
## The following object is masked from 'package:ggplot2':
##
##
      margin
library(tidyverse)
## -- Attaching packages ------ tidyverse 1.3.1 --
## v tibble 3.1.5 v dplyr 1.0.7
                   v stringr 1.4.0
## v tidyr 1.1.4
## v readr 2.0.2 v forcats 0.5.1
## v purrr 0.3.4
## -- Conflicts ----- tidyverse_conflicts() --
                  masks randomForest::c
masks stats::filter()
masks stats::lag()
masks caref
                        masks randomForest::combine()
## x dplyr::combine()
## x dplyr::filter()
## x dplyr::lag()
## x purrr::lift()
## x randomForest::margin() masks ggplot2::margin()
setwd("~/CSP571ProjectGroup")
df <- read_csv("df_with_class.csv")</pre>
## New names:
## * `` -> ...1
## Rows: 114660 Columns: 17
## -- Column specification -----
## Delimiter: ","
## chr (10): number, incident_state, sys_updated_by, contact_type, category, su...
## dbl (6): ...1, reassignment_count, reopen_count, sys_mod_count, problem_id,...
## lgl (1): made_sla
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df %>% group_by(class) %>% summarise(n = n())
## # A tibble: 19 x 2
##
     class
##
      <chr> <int>
## 1 >28days 5807
## 2 Omins
             27618
## 3 10days
             3499
## 4 10mins 2326
## 5 14days
             5356
## 6 1day
              9859
## 7 1hr
              2934
## 8 28days
              6425
## 9 2days
              8289
## 10 30mins
              4331
## 11 3days
              5244
## 12 3hr
              5316
## 13 4days
              4079
## 14 5days
              3609
## 15 5mins
              5890
## 16 6days
              3551
## 17 6hr
              3928
## 18 7days
              3617
## 19 8days
              2982
df = subset(df, select=-c(sys_updated_by, number, subcategory, resolved_updated_diff))
df = subset(df, select=-1)
df$class = as.factor(df$class)
# try on the full data set (df_small from previous version of the code)
df_small = df
train_idx = createDataPartition(df_small$class, p = 0.8, list=F)
df_train = df_small[train_idx,]
df_test = df_small[-train_idx,]
# try a randomforest model
rf_mod = randomForest(class~., data=df_train, ntree=1000)
df_test_X = subset(df_test, select=-c(class))
df_test_y = df_test$class
pred = predict(rf_mod, df_test)
columns = c("pred", "actual")
acc_df = data.frame(matrix(nrow=nrow(df_test), ncol = length(columns)))
colnames(acc_df) = columns
acc_df$pred = c(pred)
acc_df$actual = c(df_test_y)
#acc_df
acc = nrow(acc_df[acc_df$pred == acc_df$actual,])
acc/nrow(df_test)
## [1] 0.3096017
table(pred, df_test$class)
```

##

## ##	pred >28days	>28days 164	44	14	7	14days 23	1day 42	1hr 12	28days 61	51	30mins 15
##	Omins	111	4894	63	33	88	171	59	107	160	74
##	10days	0	1	2	0	0	0	0	0	0	1
##	10mins	0		0	0	0	0	0	0	0	0
##	14days	0	3	0	0	8	3	0	2	3	0
##	1day	629	458	463	130	710	1368	276	807	1041	317
##	1hr	0	0	0	0	0	0	1	0	0	0
##	28days	11	5	8	0	11	14	6	50	9	6
##	2days	19	44	31	5	47	69	12	35	113	11
##	30mins	188	21	107	244	151	259	181	179	225	368
##	3days	1		0	0	0	0	1	1	2	2
##	3hr	0		1	0	1	0	0	0	1	0
##	4days	1	2	0	0	1	1	1	1	0	0
##	5days	0	0	2	0	0	1	0	0	1	0
##	5mins	37	49	7	46	31	41	36	40	48	72
##	6days	0	0	0	0	0	0	1	1	1	0
##	6hr	0	0	0	0	0	1	0	0	0	0
##	7days	0	0	1	0	0	0	0	0	1	0
##	8days	0	0	0	0	0	1	0	1	1	0
##											
##	pred	3days	3hr 4d	avs 5da [.]	zs 5mins	6days	6hr	7days	8days		
	-	July		•		•			Juay		
##	>28days	23	21	11	18 36	17	16	20	12		
## ##	-	23		11		17	16 79	20 59	12 30		
	>28days	23 88 0	21 109 0	11	18 36	5 17 5 53	16	20	12 30		
##	>28days Omins 10days 10mins	23 88 0 0	21 109 0 0	11 74 1 0	18 36 31 216 1 (5 17 5 53 0 0	16 79 0 0	20 59 1	12 30 0 0		
## ##	>28days Omins 10days 10mins 14days	23 88 0 0 5	21 109 0 0 2	11 74 1 0 3	18 36 51 216 1 0 0 0	5 17 5 53 0 0 0 0	16 79 0 0	20 59 1 0	12 30 0 0 4		
## ## ##	>28days Omins 10days 10mins	23 88 0 0 5	21 109 0 0 2	11 74 1 0 3	18 36 31 216 1 (5 17 5 53 0 0 0 0	16 79 0 0	20 59 1	12 30 0 0 4		
## ## ## ##	>28days Omins 10days 10mins 14days	23 88 0 0 5 699	21 109 0 0 2 621	11 74 1 0 3 546 4	18 36 51 216 1 (0 (0 (31 508 0 (5 17 5 53 0 0 0 0 1 1 3 492 0 0	16 79 0 0 1 505	20 59 1 0 0 501	12 30 0 0 4 431		
## ## ## ##	>28days Omins 10days 10mins 14days 1day	23 88 0 0 5 699 0	21 109 0 0 2 621 0	11 74 1 0 3 546 4 0	18 36 31 216 1 (0 (0 (31 508 0 (6 4	5 17 5 53 0 0 0 0 1 1 3 492 0 7	16 79 0 0 1 505 0	20 59 1 0 0 501 0	12 30 0 0 4 431 0 2		
## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr	23 88 0 0 5 699 0 8 44	21 109 0 0 2 621 0 11 48	11 74 1 0 3 546 4 0 9	18 36 1 216 1 0 0 0 31 508 0 0 6 4 35 26	5 17 5 53 0 0 0 0 1 1 3 492 0 0 4 7 5 36	16 79 0 0 1 505 0 9	20 59 1 0 501 0 8 33	12 30 0 0 4 431 0 2 2 21		
## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days	23 88 0 0 5 699 0 8 44	21 109 0 0 2 621 0 11 48 202	11 74 1 0 3 546 4 0 9	18 36 31 216 1 0 0 0 31 508 0 6 4 35 26 93 275	5 17 5 53 0 0 0 0 1 1 3 492 0 0 4 7 5 36	16 79 0 0 1 505 0 9 29	200 599 1 0 501 0 8 33 79	12 30 0 0 4 431 0 2 2 3 74		
## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days	23 88 0 0 5 699 0 8 44	21 109 0 0 2 621 0 11 48 202 2	11 74 1 0 3 546 46 0 9 37 106	18 36 31 216 1 (0 0 (0 31 508 0 (6 4 35 26 93 275 0 (0	5 17 5 53 0 0 0 0 1 1 492 0 7 6 36 5 86 0 0	16 79 0 0 1 505 0 9 29 118 1	200 599 1 0 501 0 8 333 79	12 30 0 0 4 431 0 2 2 3 21 74		
## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins	23 88 0 0 5 699 0 8 44 153 4	21 109 0 0 2 621 0 11 48 202 2	11 74 1 0 3 546 4 0 9 37 106 1	18 36 31 216 1 (0 0 (0 31 508 0 (6 435 26 93 275 0 (1	5 17 5 53 0 0 0 0 1 1 3 492 0 0 4 7 6 36 86 0 0	16 79 0 0 1 505 0 9 29 118 1	200 599 1 0 501 0 8 33 79	12 30 0 0 4 431 0 2 2 3 21 74		
## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days	23 88 0 0 5 699 0 8 44 153 4 0	21 109 0 0 2 621 0 11 48 202 2 1 0	11 74 1 0 3 546 4 0 9 37 106 1 0 5	18 36 31 216 1 0 0 0 0 0 31 508 0 0 6 4 35 26 93 278 0 0 1 2 2	5 17 5 53 0 0 0 0 1 1 3 492 0 0 4 7 5 36 5 86 0 0	16 79 0 0 1 505 0 9 29 118 1 1	200 599 1 0 501 0 8 333 799 0 1	12 30 0 0 4 431 0 2 2 3 21 74 2 0		
## ###################################	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days 3hr	23 88 0 0 5 699 0 8 44 153 4 0	21 109 0 0 2 621 0 11 48 202 2 1 0	11 74 1 0 3 546 4 0 9 37 106 1 0 5	18 36 1 216 1 0 0 0 0 0 31 508 0 0 6 4 35 26 0 0 0 1 2 2 4 2	5 17 5 53 0 0 0 0 1 1 3 492 0 0 4 7 6 36 5 86 0 0 1 1 2 1	16 79 0 0 1 505 0 9 29 118 1 1 2	200 599 1 0 501 0 8 33 79 0 1 1	12 30 0 0 4 431 0 2 2 3 21 74 2 0 1 3		
## ## ## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days 3hr 4days 5days 5mins	23 88 0 0 5 699 0 8 44 153 4 0 1 0 23	21 109 0 0 2 621 0 11 48 202 2 1 0 1 44	11 74 1 0 3 546 4 0 9 37 106 1 0 5 1	18 36 31 216 1 0 0 0 31 508 0 6 435 26 335 275 0 0 1 2 2 4 220 107	5 17 5 53 0 0 0 0 1 1 3 492 0 0 4 7 5 36 5 86 0 1 1 1 2 1 1 14	16 79 0 0 1 505 0 9 29 118 1 1 2 0 22	200 599 1 0 501 0 8 333 79 0 1 1 1 0 0	12 30 0 0 4 431 0 2 2 2 74 2 0 1 3 1 4 1		
## ## ## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days 3hr 4days 5days 5mins 6days	23 88 0 0 5 699 0 8 44 153 4 0 1 0 23	21 109 0 0 2 621 0 11 48 202 2 1 0 1 1 44 0	11 74 1 0 3 546 4 0 9 37 106 1 0 5 1	18 36 31 216 1 (0 0 (0 31 508 0 (6 435 26 335 26 0 (1 2 2 2 4 22 0 (0)	5 17 5 53 0 0 0 0 1 1 492 0 0 7 6 36 6 86 0 1 1 2 1 12 1 14 0	16 79 0 0 1 505 0 9 29 118 1 1 2 0 22 1	200 599 1 0 501 0 8 333 799 0 11 0 0 15	12 30 0 0 4 431 0 2 2 3 21 74 2 0 1 3 1 4 0		
## ## ## ## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days 3hr 4days 5days 5mins 6days 6hr	23 88 0 0 5 699 0 8 44 153 4 0 1 0 23 0	21 109 0 0 2 621 0 11 48 202 2 1 0 1 1 44 0	11 74 1 0 3 546 46 0 9 37 106 1 0 5 1 21 0 0	18 36 31 216 1 (0 0 (0 831 508 0 (6 435 26 93 275 0 (1 2 2 4 2 20 107 0 (1	5 17 5 53 0 0 0 0 1 1 492 0 7 6 36 86 0 0 1 1 2 1 1 2 1 1 7 0 0 0	16 79 0 0 1 505 0 9 29 118 1 1 2 0 22 1	200 599 1 0 501 0 8 333 79 0 1 1 0 0 0 0	12 30 0 0 4 431 0 2 21 74 2 0 1 3 14 0 0		
## ## ## ## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days 3hr 4days 5days 5mins 6days 6hr 7days	23 88 0 0 5 699 0 8 44 153 4 0 1 0 23 0	21 109 0 0 2 621 0 11 48 202 2 1 0 1 1 44 0	111 74 1 0 3 546 44 0 9 37 106 1 0 5 1 21 0 0	18 36 31 216 1 (0 0 (0 31 508 0 (6 435 26 335 26 0 (1 2 2 2 4 2 20 107 0 (1 0 (1 0 (1 0 (1) 0 (1 0 (1) 0 (1 0 (1) 0	5 17 5 53 0 0 0 0 1 1 492 0 0 4 7 6 36 86 0 0 1 1 1 2 1 1 1 2 1 1 0 0 0 0	16 79 0 0 1 505 0 9 29 118 1 1 2 0 22 1	200 599 11 00 501 0 8 333 799 0 11 10 0 0 15 0 0 0 0 0 0 0 0 0 0 0 0	12 30 0 0 4 431 0 2 21 74 2 0 1 3 14 0 0		
## ## ## ## ## ## ## ## ##	>28days Omins 10days 10mins 14days 1day 1hr 28days 2days 30mins 3days 3hr 4days 5days 5mins 6days 6hr	23 88 0 0 5 699 0 8 44 153 4 0 1 0 23 0	21 109 0 0 2 621 0 11 48 202 2 1 0 1 1 44 0	11 74 1 0 3 546 46 0 9 37 106 1 0 5 1 21 0 0	18 36 31 216 1 (0 0 (0 831 508 0 (6 435 26 93 275 0 (1 2 2 4 2 20 107 0 (1	5 17 5 53 0 0 0 0 1 1 492 0 0 4 7 6 36 86 0 0 1 1 1 2 1 1 1 2 1 1 0 0 0 0	16 79 0 0 1 505 0 9 29 118 1 1 2 0 22 1	200 599 1 0 501 0 8 333 79 0 1 1 0 0 0 0	12 30 0 0 4 431 0 2 2 3 21 74 2 0 1 3 14 0 0		

summary(rf_mod)

```
##
                   Length Class Mode
## call
                         4 -none- call
## type
                         1 -none- character
## predicted
                     91737 factor numeric
                     20000 -none- numeric
## err.rate
## confusion
                       380 -none- numeric
## votes
                   1743003 matrix numeric
## oob.times
                     91737 -none- numeric
## classes
                       19 -none- character
## importance
                       11 -none- numeric
## importanceSD
                        O -none- NULL
```

```
O -none- NULL
## localImportance
## proximity
                    O -none- NULL
## ntree
                    1 -none- numeric
## mtry
                    1 -none- numeric
## forest
## y
                   14 -none- list
                91737 factor numeric
## test
                    O -none- NULL
## inbag
                    O -none- NULL
## terms
                     3 terms call
```