Patient No-show Prediction Model

```r
set.seed(10-27-25)

library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.5.2
## v ggplot2   4.0.0     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(randomForest)
```

```
## randomForest 4.7-1.2
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
##
## The following object is masked from 'package:dplyr':
##
##     combine
##
## The following object is masked from 'package:ggplot2':
##
##     margin
```

Create Model

Read in training and testing dataset

```r
train_file <- "train_dataset.csv.gz"
test_file <- "test_dataset.csv.gz"

train_raw <- read_csv(train_file, guess_max = 10000)
```

```
## Rows: 36588 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): id, provider_id, address, age, specialty, no_show
## dttm (1): appt_time
## date (1): appt_made
##
## 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.
```

```r
test_raw <- read_csv(test_file, guess_max = 10000)
```

```
## Rows: 36631 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (6): id, provider_id, address, age, specialty, no_show
## dttm (1): appt_time
## date (1): appt_made
##
## 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.
```

Feature Engineering

```r
# Parse time
train <- train_raw %>%
mutate(
appt_time = ymd_hms(appt_time, tz = "UTC"),
appt_made = as_date(appt_made)
)

test <- test_raw %>%
mutate(
appt_time = ymd_hms(appt_time, tz = "UTC"),
appt_made = as_date(appt_made)
)

# create lead_time_days, appt hour/day, weekend flag

train <- train %>%
mutate(
lead_time_days = as.numeric(difftime(appt_time, appt_made, units = "days")),
appt_hour = hour(appt_time),
appt_wday = wday(appt_time, label = TRUE, week_start = 1), # Monday = 1
is_weekend = if_else(appt_wday %in% c("Sat", "Sun"), 1, 0)
)

test <- test %>%
mutate(
lead_time_days = as.numeric(difftime(appt_time, appt_made, units = "days")),
appt_hour = hour(appt_time),
appt_wday = wday(appt_time, label = TRUE, week_start = 1),
is_weekend = if_else(appt_wday %in% c("Sat", "Sun"), 1, 0)
)

# make no show categorical

train$no_show <- as.factor(train$no_show)

test$no_show <- as.factor(test$no_show)
```

Modeling

```
features <- c("age", "address", "specialty", "provider_id",
"lead_time_days", "appt_hour", "is_weekend")

formula <- as.formula(paste("no_show ~", paste(features, collapse = " + ")))

rf_model <- randomForest(
  formula,
  data = train,
  ntree = 200,
  importance = TRUE
)
```

Model Results

```
print(rf_model)
```

```
##
## Call:
##  randomForest(formula = formula, data = train, ntree = 200, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 200
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 11.54%
## Confusion matrix:
##       0     1 class.error
## 0 21102  1881   0.0818431
## 1  2341 11264   0.1720691
```

```
# Class predictions
rf_pred_class <- predict(rf_model, newdata = test, type = "response")

# Class probabilities
rf_pred_prob <- predict(rf_model, newdata = test, type = "prob")
```

```
# True values
actual <- test$no_show

# Error rate = misclassified / total
error_rate <- mean(rf_pred_class != actual)

cat("Overall error rate:", round(error_rate, 4))
```

```
## Overall error rate: 0.1123
```