FDA Drug Recall Classifier - Modeling Report

Jannet Castaneda

Load and Prepare Data

```
fda <- read_csv("2_data/fda_recalls_clean.csv")

# Ensure classification is a factor
fda <- fda %>%
  mutate(event_classification = as.factor(event_classification))
```

Split Data

```
set.seed(123)
split <- initial_split(fda, prop = 0.8)
train_data <- training(split)
test_data  <- testing(split)
```

Recipe for Text Features

```
recall_recipe <- recipe(event_classification ~ reason_for_recall, data = train_data) %>%
  step_tokenize(reason_for_recall) %>%
  step_stopwords(reason_for_recall) %>%
  step_tokenfilter(reason_for_recall, max_tokens = 100) %>%
  step_tfidf(reason_for_recall)
```

Model Specification and Workflow

```
lr_spec <- multinom_reg() %>%
  set_engine("nnet") %>%
  set_mode("classification")

recall_wf <- workflow() %>%
  add_recipe(recall_recipe) %>%
  add_model(lr_spec)
```

Fit the Model

```
recall_fit <- recall_wf %>%
  fit(data = train_data)
```

Evaluate on the Test Set

```
test_predictions <- predict(recall_fit, test_data, type = "prob") %>%
  bind_cols(predict(recall_fit, test_data)) %>%
  bind_cols(test_data)
```

Confusion Matrix

```
conf_mat(test_predictions, truth = event_classification, estimate = .pred_class)
```

```
##            Truth
## Prediction  Class I Class II Class III
##   Class I       358       83        10
##   Class II      112     2438       166
##   Class III       7       93       115
```

Accuracy and Kappa

```
metrics(test_predictions, truth = event_classification, estimate = .pred_class)
```

```
## # A tibble: 2 x 3
##   .metric  .estimator .estimate
##   <chr>    <chr>          <dbl>
## 1 accuracy multiclass     0.861
## 2 kap      multiclass     0.608
```

Manual Keyword Inspection: "death"

```
# Count how often each classification is assigned for reasons containing 'death'
fda %>%
  filter(str_detect(tolower(reason_for_recall), "death")) %>%
  count(event_classification) %>%
  arrange(desc(n))
```

```
## # A tibble: 1 x 2
##   event_classification     n
##   <fct>                <int>
## 1 Class I                  1
```

```r
# View examples of these recall reasons
fda %>%
  filter(str_detect(tolower(reason_for_recall), "death")) %>%
  select(reason_for_recall, event_classification) %>%
  head(10)
```

```
## # A tibble: 1 x 2
##   reason_for_recall                                          event_classification
##   <chr>                                                      <fct>
## 1 Marketed Without An Approved NDA/ANDA: FDA analysis foun~ Class I
```

Save Trained Model

```r
dir.create("4_model-api", showWarnings = FALSE)
saveRDS(recall_fit, "4_model-api/recall_model.rds")
```