

Model Selection Results

Eduardo Yuki Yada

```
library(tidyverse)
library(kableExtra)

niceFormatting = function(df, caption="", digits = 2, font_size = NULL, label = 1){
  df %>%
    kbl(booktabs = T, longtable = T, caption = caption, digits = digits, format = "latex", label = label) %>%
    kable_styling(font_size = font_size,
                  latex_options = c("striped", "HOLD_position", "repeat_header"))
}

path <- "./auxiliar/model_selection/performance"

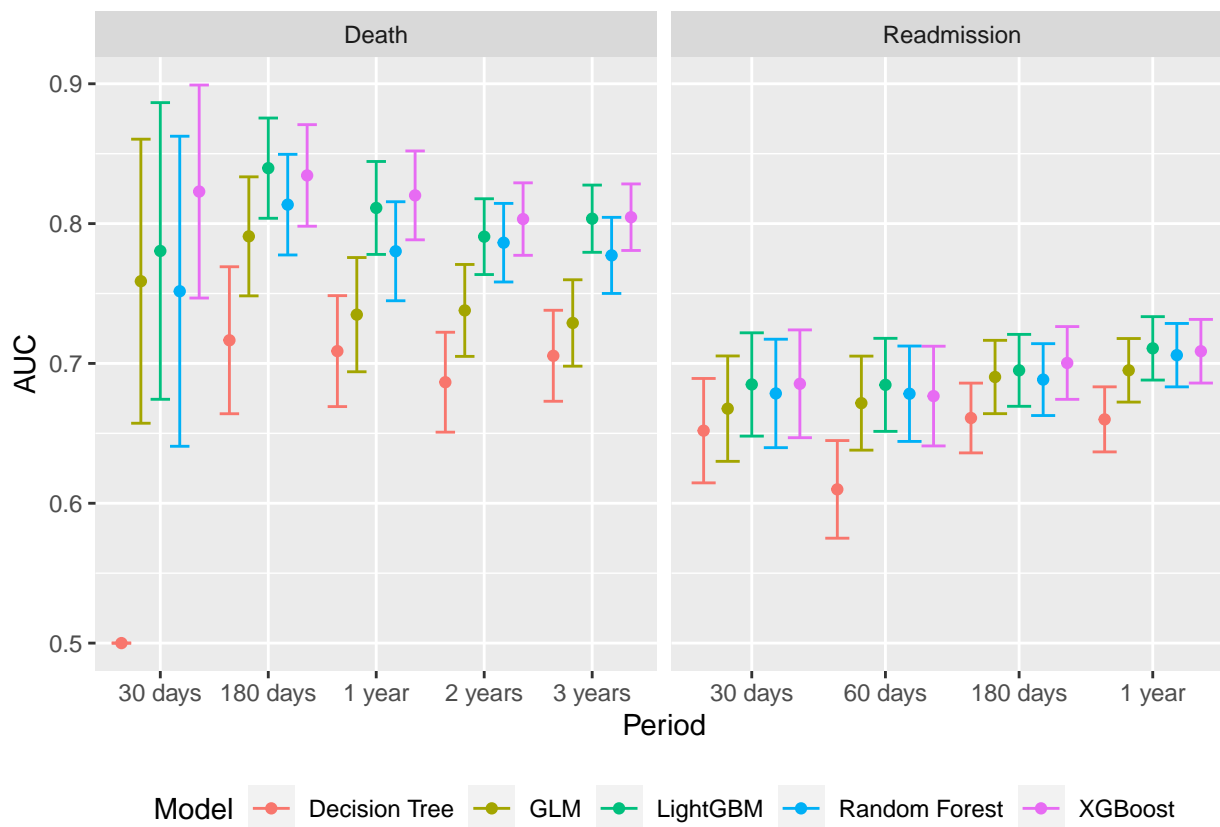
df <- tibble()

for (file in list.files(path)) {
  df <- df %>% bind_rows(readRDS(file.path(path, file)))
}

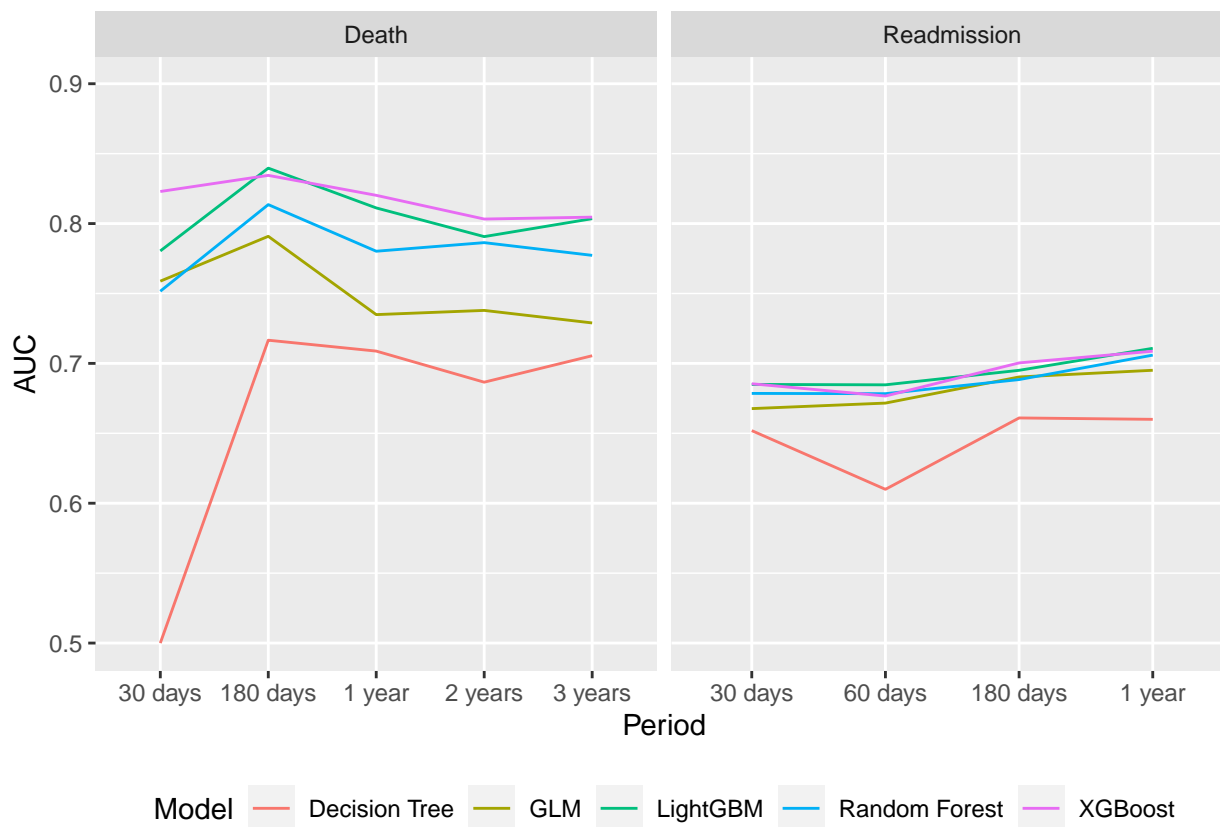
df <- df %>%
  mutate(Period = gsub(".*_", "", Target),
         Target = gsub("_.*", "", Target) %>% str_to_title,
         Period = factor(
           Period,
           levels = c(
             '30d',
             '30days',
             '60d',
             '180d',
             '180days',
             '1year',
             '2year',
             '3year'
           ),
           labels = c(
             '30 days',
             '30 days',
             '60 days',
             '180 days',
             '180 days',
             '1 year',
             '2 years',
             '3 years'
           )
         ))

df %>%
  ggplot(aes(
    x = Period,
    y = AUC,
    ymin = `Lower Limit`,
    ymax = `Upper Limit`,
    color = Model
```

```
)) +
facet_grid(~Target, scales = "free_x") +
geom_point(position = position_dodge(width = 0.9)) +
geom_errorbar(position = position_dodge(width = 0.9)) +
theme(legend.position = "bottom")
```



```
df %>%
ggplot(
  aes(
    x = Period,
    y = AUC,
    ymin = `Lower Limit`,
    ymax = `Upper Limit`,
    group = Model,
    colour = Model
  )
) +
facet_grid(~Target, scales = "free_x") +
geom_line() +
theme(legend.position = "bottom")
```



```
df %>%
  filter(Target == 'Readmission') %>%
  select(Period, Model, AUC, `Lower Limit`, `Upper Limit`) %>%
  niceFormatting(caption = 'Readmission models comparison', label = 1)
```

Table 1: Readmission models comparison

Period	Model	AUC	Lower Limit	Upper Limit
180 days	XGBoost	0.70	0.67	0.73
180 days	LightGBM	0.70	0.67	0.72
180 days	GLM	0.69	0.66	0.72
180 days	Decision Tree	0.66	0.64	0.69
180 days	Random Forest	0.69	0.66	0.71
1 year	XGBoost	0.71	0.69	0.73
1 year	LightGBM	0.71	0.69	0.73
1 year	GLM	0.70	0.67	0.72
1 year	Decision Tree	0.66	0.64	0.68
1 year	Random Forest	0.71	0.68	0.73
30 days	XGBoost	0.69	0.65	0.72
30 days	LightGBM	0.68	0.65	0.72
30 days	GLM	0.67	0.63	0.71
30 days	Decision Tree	0.65	0.61	0.69
30 days	Random Forest	0.68	0.64	0.72
60 days	XGBoost	0.68	0.64	0.71
60 days	LightGBM	0.68	0.65	0.72
60 days	GLM	0.67	0.64	0.71
60 days	Decision Tree	0.61	0.58	0.64
60 days	Random Forest	0.68	0.64	0.71

```
df %>%
  filter(Target == 'Death') %>%
```

```
select(Period, Model, AUC, `Lower Limit`, `Upper Limit`) %>%
  niceFormatting(caption = 'Death models comparison', label = 2)
```

Table 2: Death models comparison

Period	Model	AUC	Lower Limit	Upper Limit
180 days	XGBoost	0.83	0.80	0.87
180 days	LightGBM	0.84	0.80	0.88
180 days	GLM	0.79	0.75	0.83
180 days	Decision Tree	0.72	0.66	0.77
180 days	Random Forest	0.81	0.78	0.85
1 year	XGBoost	0.82	0.79	0.85
1 year	LightGBM	0.81	0.78	0.84
1 year	GLM	0.73	0.69	0.78
1 year	Decision Tree	0.71	0.67	0.75
1 year	Random Forest	0.78	0.74	0.82
2 years	XGBoost	0.80	0.78	0.83
2 years	LightGBM	0.79	0.76	0.82
2 years	GLM	0.74	0.71	0.77
2 years	Decision Tree	0.69	0.65	0.72
2 years	Random Forest	0.79	0.76	0.81
30 days	XGBoost	0.82	0.75	0.90
30 days	LightGBM	0.78	0.67	0.89
30 days	GLM	0.76	0.66	0.86
30 days	Decision Tree	0.50	0.50	0.50
30 days	Random Forest	0.75	0.64	0.86
3 years	XGBoost	0.80	0.78	0.83
3 years	LightGBM	0.80	0.78	0.83
3 years	GLM	0.73	0.70	0.76
3 years	Decision Tree	0.71	0.67	0.74
3 years	Random Forest	0.78	0.75	0.80

```
path <- "./auxiliar/model_selection/performance"

df <- tibble()

for (file in list.files(path)) {
  df <- df %>% bind_rows(readRDS(file.path(path, file)))
}

df <- df %>%
  mutate(Period = gsub(".*_", "", Target),
         Target = gsub("_.*", "", Target) %>% str_to_title,
         Period = factor(
           Period,
           levels = c(
             '30d',
             '30days',
             '60d',
             '180d',
             '180days',
             '1year',
             '2year',
             '3year'
           ),
           labels = c(
             '30 days',
             '30 days',

```

```

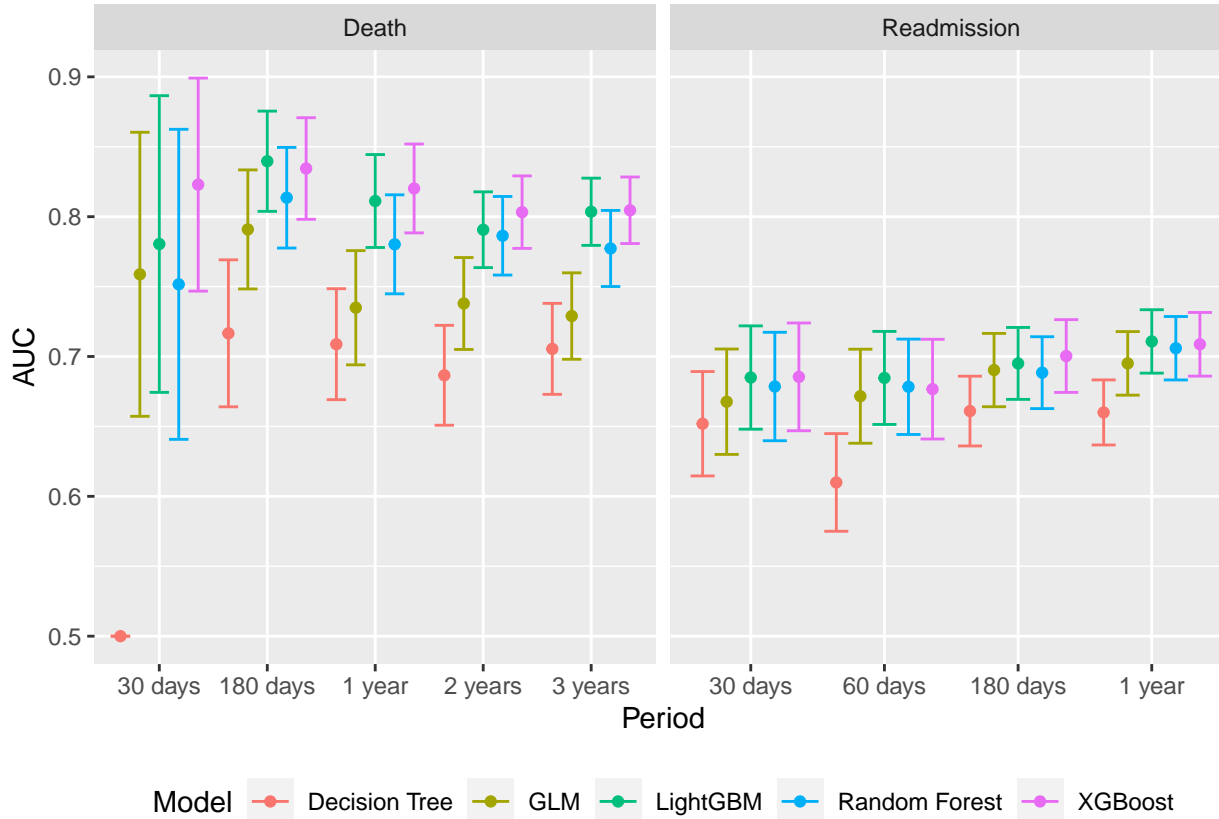
'60 days',
'180 days',
'180 days',
'1 year',
'2 years',
'3 years'
)
))

```

```

df %>%
  ggplot(aes(
    x = Period,
    y = AUC,
    ymin = `Lower Limit`,
    ymax = `Upper Limit`,
    color = Model
  )) +
  facet_grid(~Target, scales = "free_x") +
  geom_point(position = position_dodge(width = 0.9)) +
  geom_errorbar(position = position_dodge(width = 0.9)) +
  theme(legend.position = "bottom")

```

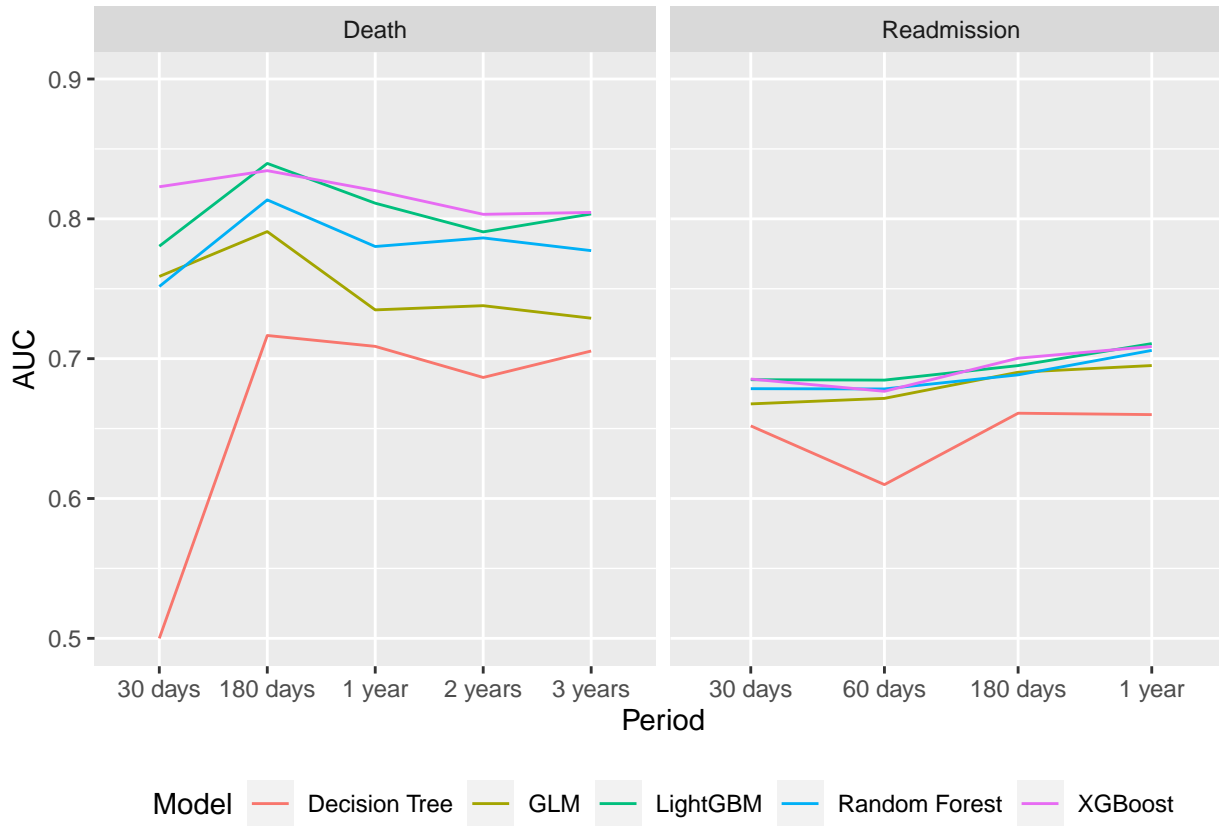


```

df %>%
  ggplot(
    aes(
      x = Period,
      y = AUC,
      ymin = `Lower Limit`,
      ymax = `Upper Limit`,
      group = Model,
      colour = Model
    )
  ) +

```

```
facet_grid(~Target, scales = "free_x") +
geom_line() +
theme(legend.position = "bottom")
```



```
df %>%
  filter(Target == 'Readmission') %>%
  select(Period, Model, AUC, `Lower Limit`, `Upper Limit`) %>%
  niceFormatting(caption = 'Readmission models comparison', label = 3)
```

Table 3: Readmission models comparison

Period	Model	AUC	Lower Limit	Upper Limit
180 days	XGBoost	0.70	0.67	0.73
180 days	LightGBM	0.70	0.67	0.72
180 days	GLM	0.69	0.66	0.72
180 days	Decision Tree	0.66	0.64	0.69
180 days	Random Forest	0.69	0.66	0.71
1 year	XGBoost	0.71	0.69	0.73
1 year	LightGBM	0.71	0.69	0.73
1 year	GLM	0.70	0.67	0.72
1 year	Decision Tree	0.66	0.64	0.68
1 year	Random Forest	0.71	0.68	0.73
30 days	XGBoost	0.69	0.65	0.72
30 days	LightGBM	0.68	0.65	0.72
30 days	GLM	0.67	0.63	0.71
30 days	Decision Tree	0.65	0.61	0.69
30 days	Random Forest	0.68	0.64	0.72
60 days	XGBoost	0.68	0.64	0.71
60 days	LightGBM	0.68	0.65	0.72
60 days	GLM	0.67	0.64	0.71
60 days	Decision Tree	0.61	0.58	0.64

Table 3: Readmission models comparison (*continued*)

Period	Model	AUC	Lower Limit	Upper Limit
60 days	Random Forest	0.68	0.64	0.71

```
df %>%
  filter(Target == 'Death') %>%
  select(Period, Model, AUC, `Lower Limit`, `Upper Limit`) %>%
  niceFormatting(caption = 'Death models comparison', label = 4)
```

Table 4: Death models comparison

Period	Model	AUC	Lower Limit	Upper Limit
180 days	XGBoost	0.83	0.80	0.87
180 days	LightGBM	0.84	0.80	0.88
180 days	GLM	0.79	0.75	0.83
180 days	Decision Tree	0.72	0.66	0.77
180 days	Random Forest	0.81	0.78	0.85
1 year	XGBoost	0.82	0.79	0.85
1 year	LightGBM	0.81	0.78	0.84
1 year	GLM	0.73	0.69	0.78
1 year	Decision Tree	0.71	0.67	0.75
1 year	Random Forest	0.78	0.74	0.82
2 years	XGBoost	0.80	0.78	0.83
2 years	LightGBM	0.79	0.76	0.82
2 years	GLM	0.74	0.71	0.77
2 years	Decision Tree	0.69	0.65	0.72
2 years	Random Forest	0.79	0.76	0.81
30 days	XGBoost	0.82	0.75	0.90
30 days	LightGBM	0.78	0.67	0.89
30 days	GLM	0.76	0.66	0.86
30 days	Decision Tree	0.50	0.50	0.50
30 days	Random Forest	0.75	0.64	0.86
3 years	XGBoost	0.80	0.78	0.83
3 years	LightGBM	0.80	0.78	0.83
3 years	GLM	0.73	0.70	0.76
3 years	Decision Tree	0.71	0.67	0.74
3 years	Random Forest	0.78	0.75	0.80