

Interpretable Machine Learning PD4

Daniel Ponikowski

4 kwietnia 2019

Wybrane zmienne :

1. ppwork - aktualny status zatrudnienia
2. w6_q20 - czy obecnie mieszkasz z partnerem?
3. Q21A_Year - w którym roku pierwszy raz spotkałeś partnera?
4. ppage - wiek

Wczytanie modelu i danych:

```
data <- read.dta13(file = "../PD1/HCMST 2017 fresh sample for public sharing draft v1.1.dta")
df <- data[,c("S1", "ppwork", "w6_q19", "Q21A_Year", "ppage")]

df <- df %>% mutate(Q21A_Year = as.numeric(as.character(Q21A_Year))
                    , ppwork = factor(ppwork)
                    , w6_q19 = factor(w6_q19)
                    , ppage = as.numeric(ppage)
                    , S1 = factor(S1)) %>%
  na.omit() %>% unique() %>% as.data.frame()
row.names(df) <- 1:nrow(df)
RF <- readRDS("../PD3/randomForestPD1.rds")
```

Wybór obserwacji:

```
(obs <- df[130,])
```

```
##           S1           ppwork w6_q19 Q21A_Year ppage
## 130 Yes, I am Married Not working - other    Yes    2010    30
```

Predykcja black-boxa dla naszej obserwacji:

```
predict(RF, obs, "prob")
```

```
##      Yes, I am Married No, I am not Married
## 130              0.994              0.006
```

LIME

Zmienne wybrane do perturbacji

1. Q21A_Year - w którym roku pierwszy raz spotkałeś partnera?
2. ppage - wiek

Do wazenia wykorzystam normę l^1 . Czyli sume wartości bezwzględnych.

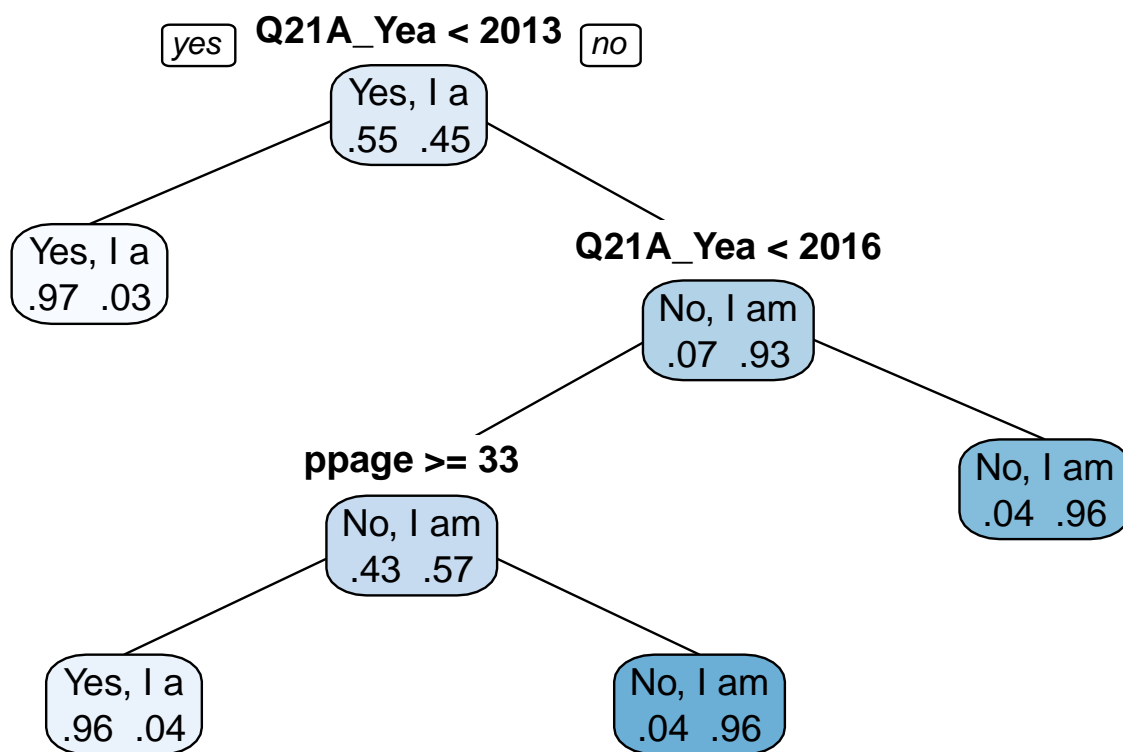
```
set.seed(2)
z <- obs[, -1]
Q21A_Year_norm <- rnorm(n = 1000, mean = z$Q21A_Year, sd = sd(df$Q21A_Year)) %>% floor()
ppage_norm <- rnorm(n = 1000, mean = z$ppage, sd = sd(df$ppage)) %>% floor()
```

```

z1 <- data.frame(ppwork = z$ppwork,
                 w6_q19 = z$w6_q19,
                 Q21A_Year = Q21A_Year_norm,
                 ppage = ppage_norm)

z1 <- rbind(z,z1)
z1 <- z1[-1,]
dist <- abs(z1$Q21A_Year - z$Q21A_Year) + abs(z1$ppage - z$ppage)
z1 <- z1 %>% mutate(S1 = factor(predict(RF,z1)))
tree <- train(S1~.,data = z1,method = "rpart2",weights = dist,
              tuneGrid = expand.grid(maxdepth = c(3)) )
prp(tree$finalModel, box.palette = "Blues", tweak = 1.2,type=1,extra = 4)

```



Wnioski:

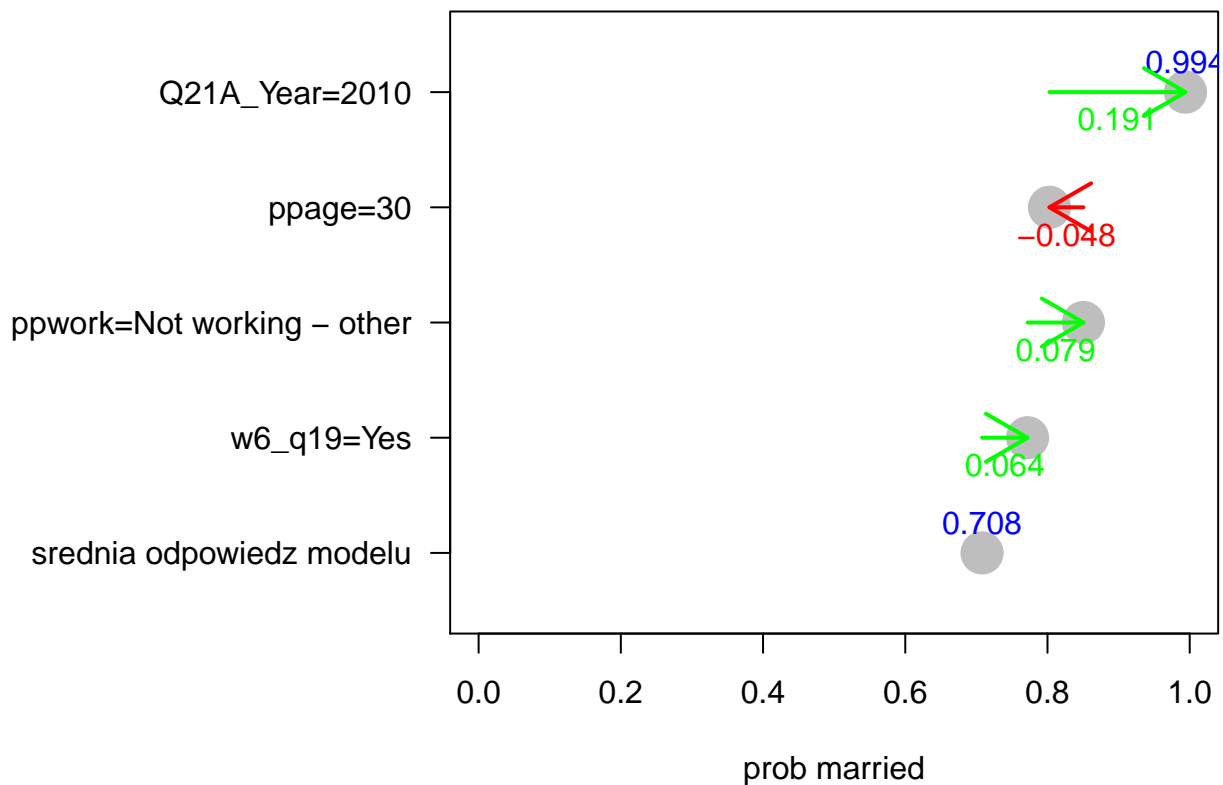
Wartosc zmiennej *Q21A_Year* według naszej białej skrzynki jest ważna, drzewo na początku dokonuje podziału względem tej zmiennej. Nasza obserwacja ma wartość *Q21A_Year* = 2010, więc jest klasyfikowana do najwyższego lewego liścia, który ma największe prawdopodobieństwo klasyfikacji do klasy osób w związku małżeńskim. Możemy więc sadzić, że ta wartość zmiennej ma największy wpływ (z badanych zmiennych) na klasyfikację naszej obserwacji. Podział po zmiennej *ppage* jest dokonywany dopiero na trzecim poziomie, co sugeruje mały wpływ wartości tej zmiennej na odpowiedź modelu.

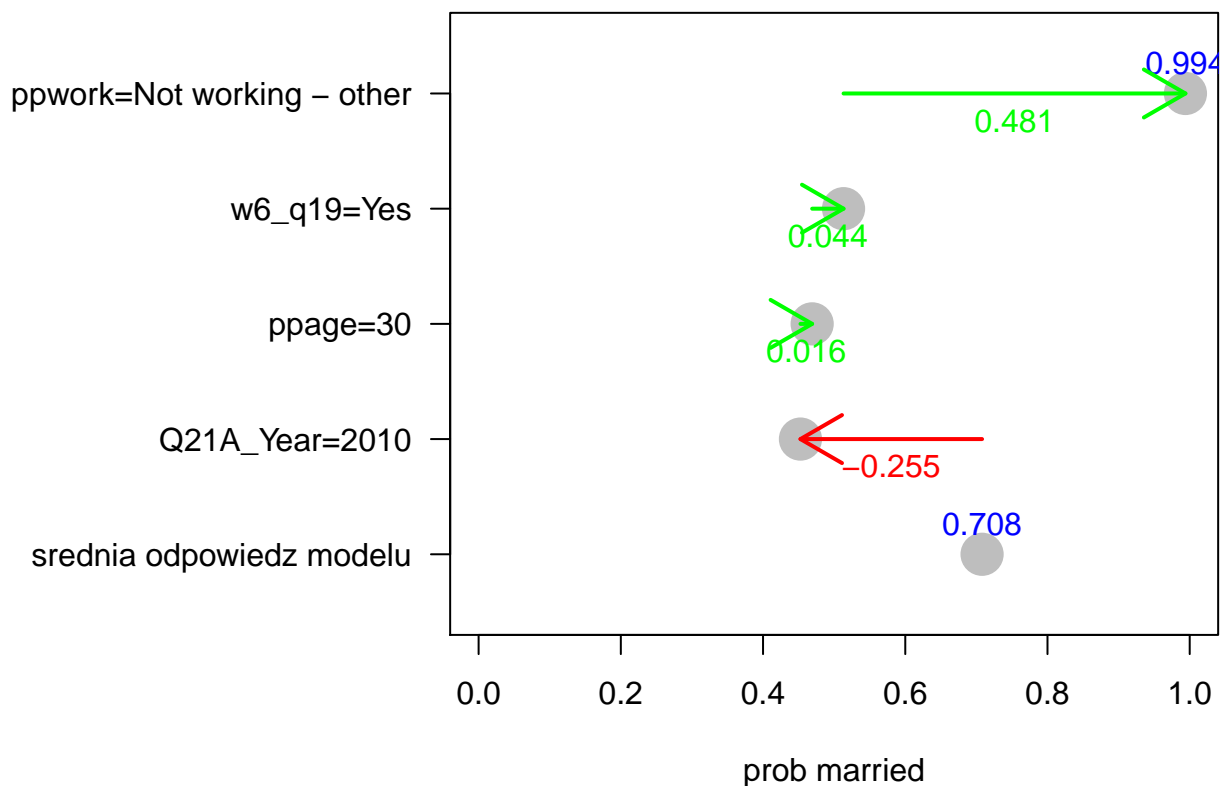
Porównanie z PD3 (break down):

```

break_down <- function(zmienne,df,ind_osoba,model){
  wynik <- data.frame(matrix(nrow = 1,ncol = 0))
  osoba <- df[ind_osoba,]
  odp_modelu <- predict(object = model,df,type = "prob")[,1] %>% mean() %>% round(3)
  wynik$odp_modelu <- odp_modelu
  for (i in 1:length(zmienne) ){
    df[[zmienne[i]]] <- osoba[[zmienne[i]]]
    wynik[[zmienne[i]]] <- predict(object = model,newdata = df,type = "prob")[,1] %>% mean()
    colnames(wynik) <- c("srednia odpowiedz modelu",zmienne)
    wynik <- wynik %>% unlist() %>% c()
    {par(oma = c(1.1,8.2,0,0))
    par(mar = c(5.1,4.2,0.1,0.1))
    plot(x = wynik, y = 1:length(wynik),pch = 16, col = "grey", cex = 3,las = 1
        ,frame.plot = TRUE,axes = FALSE,xlab = "prob married"
        ,ylab = "",ylim = c(0.5,length(wynik)+0.5),xlim = c(0,1))
    axis(side = 1,at = seq(0,1,length.out = 6))
    axis(side = 2,at = 1:length(wynik),labels = c("srednia odpowiedz modelu",paste0(
      zmienne,"=",lapply(osoba,FUN = as.character) %>% unlist %>% "["(zmienne))),las = 1)
    text(wynik[1],x = wynik[1],y = 1,pos = 3,col = "blue")
    text(wynik[length(wynik)],y = length(wynik),x = wynik[length(wynik)],pos = 3,col = "blue")
    for (i in 1:(length(wynik)-1)){
      roznica <- round(wynik[i+1]-wynik[i],3) %>% as.character()
      col <- ifelse(as.numeric(roznica) > 0,"green","red")
      arrows(y0 = i+1,y1 = i+1,x0 = wynik[i],x1 = wynik[i+1],lwd = 2,col = col)
      text(roznica ,x = (wynik[i+1] + wynik[i])/2 ,y = i + 1,pos = 1,col = col)}}}

```





Wnioski:

W zależności od kolejności zmiennych, wartość zmiennej *Q21A_Year* wpływa pozytywnie bądź negatywnie na klasyfikację do klasy osób w związku małżeńskim. Jednak dla obu przedstawionych kolejności, wartość tej zmiennej mocno zmienia średnią odpowiedź modelu. Więc interpretacja ta pokrywa się z interpretacją uzyskaną dzięki metodzie LIME (wartość tej zmiennej mocno wpływa na predykcję modelu). Zmienna *ppage* wydaje się być mało wpływająca na odpowiedź modelu dla naszej obserwacji, podobne wnioski możemy wysnuć z metody LIME, gdzie stworzone drzewo dokonuje podziału względem tej zmiennej dopiero na głębokości 3.