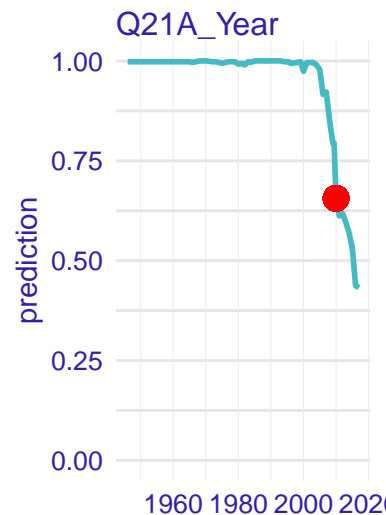
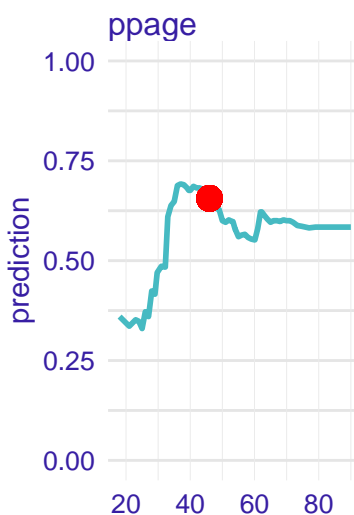
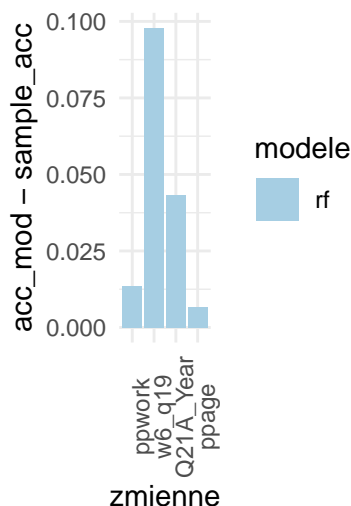


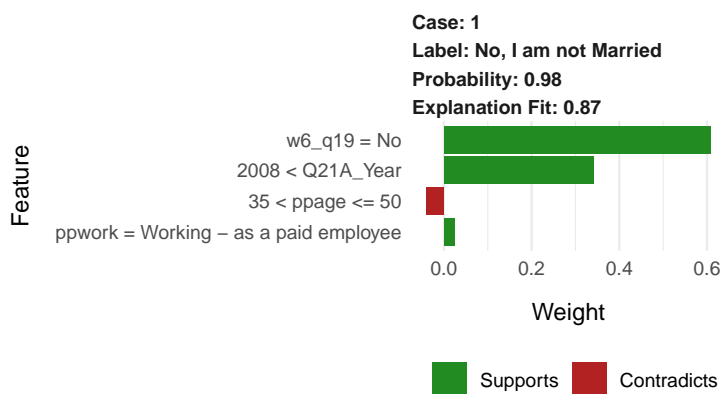
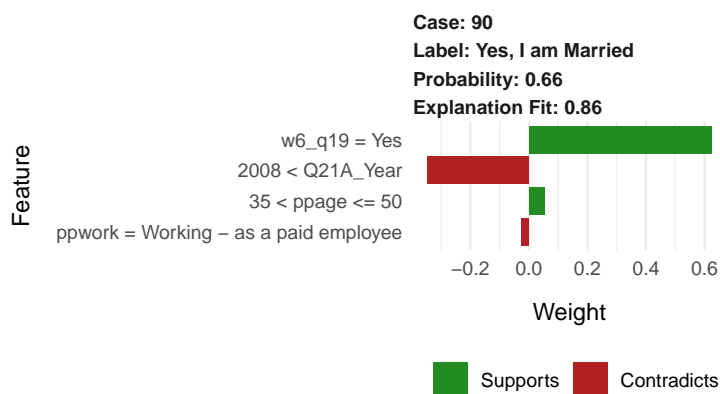
PD10 Interpretable ML - Daniel Ponikowski

W tym opracowaniu użyję zmiennych (**Q21A_Year**, **w6_q19**, **ppage**, **ppwork**), oraz modelu (**las losowy**), których używałem w poprzednich pracach domowych. Spróbuję wyjaśnić przypadek w którym las się myli (dokładnie dla osoby o następujących wartościach zmiennych: **Q21A_Year** : 2010, **w6_q19** : tak, **ppage** : 46, **ppwork** : praca - jako opłacany pracownik). Jest to osoba, która w rzeczywistości nie jest w związku małżeńskim, a model przewiduje jej związek małżeński. Na początku sprawdzę ważność zmiennych użytych w modelu (pierwszy wykres poniżej) - najważniejsza okazała się nominalna zmienna, czyli odpowiedź na pytanie "czy mieszkasz z partnerem?". W dalszej części sprawdzę czy tym przypadkiem okazała się ona decydująca. Do analizy zmiennych ilościowych zastosowałem poniższe wykresy CP odczytujemy, że poprzez zmianę wartości jednej z tych zmiennych możemy otrzymać inną odpowiedź modelu (np. wystarczy że osoba byłaby młodsza).

Variable importance



Dla zmiennych jakościowych, bardzo mocno odpowiedź modelu "podbija" wartość zmiennej **w6_q19**. Mieszkanie z partnerem jest rzeczą "standardową" dla małżeństw. Opisana powyżej własność tej zmiennej pokazuje poniższy wykres LIME (czyli lokalnego przybliżenia lasu losowego). Drugą bardzo ważną zmienną dla odpowiedzi modelu w pobliżu analizowanej obserwacji jest zmienna **Q21A_Year**, czyli rok poznania obecnego partnera. Nasza obserwacja poznała obecnego partnera w 2010 roku, czyli stosunkowo niedawno, obniża to odpowiedź modelu (co jest zgodne z moją intuicją, w związkach małżeńskich zazwyczaj są osoby długo znające się).



Wnioski:

1. Dlaczego? Odpowiedź modelu jest w dużym stopniu spowodowana zmienną **w6_q19**, oraz wiekiem obserwacji
2. Co zrobić żeby podnieść wynik modelu? Żeby zwiększyć przewidywane prawdopodobieństwo małżeństwa, trzeba zmniejszyć rok poznania się osób. Dla zmniejszenia prawdopodobieństwa, najlepiej zmienić wartość zmiennej **w6_q19** (prawy wykres LIME).
3. Jak wyglądały wyniki dla podobnych osób? W zależności od zmiany wartości zmiennych ciągłych predykcja prawdopodobieństwa małżeństwa może się wahać od ok 30% do 100%. Jeżeli chodzi o zmienne kategoryczne to zmienna **w6_q19** w znaczący sposób wpływa na predykcję modelu (dwa wykresy LIME- lewy dla oryginalnej zmiennej, prawy dla obserwacji ze zmienioną wartością **w6_q19**), wartość **No** mocno zwiększa prawdopodobieństwo nie bycia w związku małżeńskim, a **Yes** w zdecydowany sposób podwyższa prawdopodobieństwo małżeństwa. Wpływ tej zmiennej dobrze oddaje różnica w predykcji dla wybranej obserwacji i dla obserwacji ze zmienioną wartością tej zmiennej (zmiana prawdopodobieństwa małżeństwa z 0.66 do 0.02)