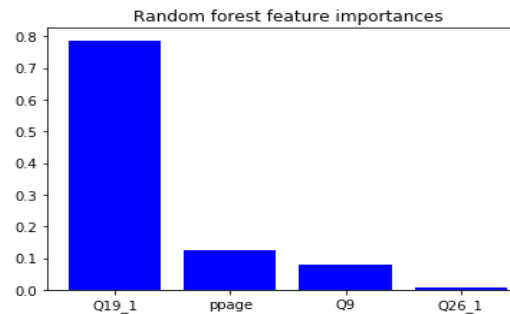
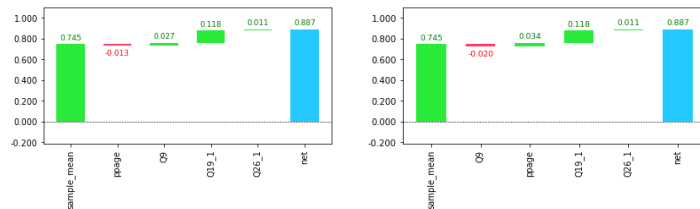


Czy Joe i Clara są małżeństwem?

Jesteście świeżo po przeprowadzce. Podczas poznawania okolicy spotykacie Joe, sąsiada mieszkającego w domu obok. Z rozmowy wynika, że Joe jest 41-latkami mieszkającym z rok młodszą partnerką Clary. Para poznała się w czasach studenckich, chociaż każde z nich studiowało na różnych uczelniach. Czy ta informacja wystarcza do wywnioskowania czy Joe i Clara są małżeństwem?

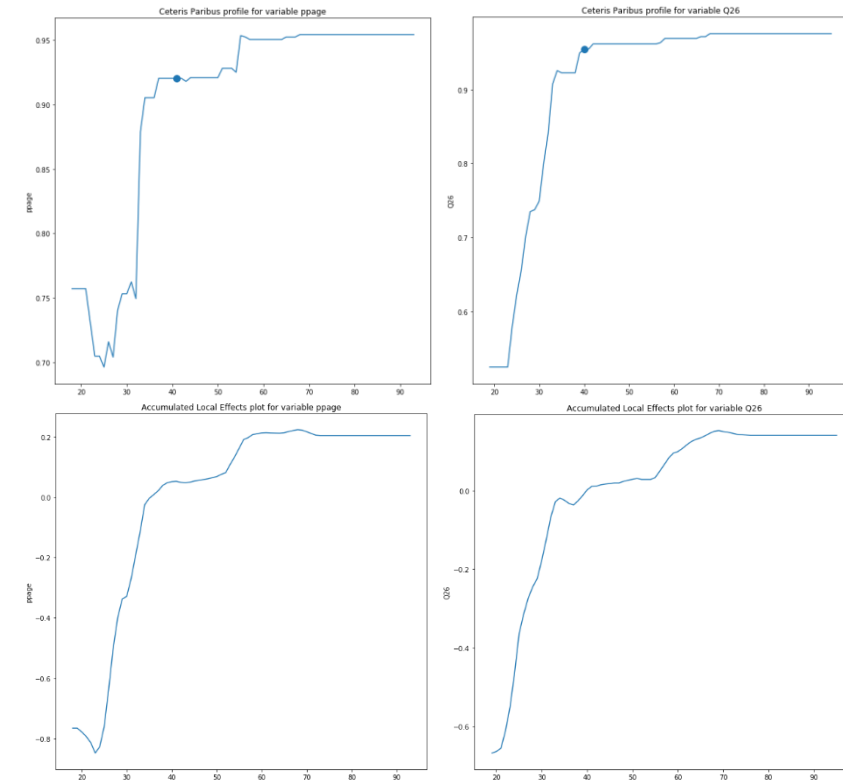
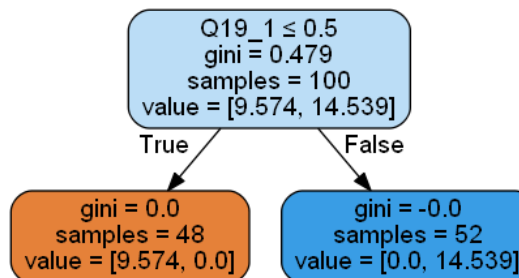
W celu odpowiedzi na to pytanie zbudowany został model predykcyjny. Poprawnie sklasyfikował on Joe i Clary jako małżeństwo. W jaki sposób klasyfikator doszedł do właściwej konkluzji?



Powyższe wykresy informują o zmianie prawdopodobieństwa bycia małżeństwem pod wpływem dodawania kolejnych informacji o Joe. Ciekawostką jest, że **wiek partnerów ma najpierw ujemny a potem dodatni wpływ** na ostateczną prognozę. Możliwe, że zmienne te wchodziły w interakcję. Są one również silnie skorelowane, co widzać na poniższej macierzy korelacji.

	ppage	Q26
ppage	1.000000	0.924767
Q26	0.924767	1.000000

Okazuje się, że model dokonuje klasyfikacji par podobnych do Joe i Clary **na podstawie informacji o mieszkaniu razem**. Wynika to z lokalnej metody LIME oraz globalnych ważności przypisanych cechom przez las losowy.



Powyższe wykresy PDP oraz ALE pokazują wpływ wieku Joe oraz Clary na prawdopodobieństwo bycia małżeństwem. Pierwsza para wykresów jest warunkową wartością oczekiwaną odpowiedzi modelu pod warunkiem wieku, druga para wykresów ukazuje zdekorowany i oczyszczony wpływ wieku na odpowiedź modelu. Trajektorie są zgodne z intuicją:

im starszy któryś z partnerów, tym większa szansa na klasyfikację pary jako małżeństwo.

Jak można poprawić model?

- Wykresy typu breakdown sugerują istnienie interakcji pomiędzy Wiekami partnerów. Sensownym rozwiązaniem byłaby próba uwzględnienia interakcji pomiędzy zmiennymi np. poprzez **zamianę jednej ze zmiennych na różnicę wieku pomiędzy partnerami**.
- Najważniejszym predyktorem jest zmienna informująca o wspólnym mieszkaniu partnerów. Kształcenie się na jednej uczelni ma marginalny wpływ na predykcję. Dobrym pomysłem byłaby **podmiana zmiennej Q26 na inną występującą w zbiorze danych**.