

27 Kasım 2023

# Açıklanabilir Yapay Zeka

## 8. Hafta: Global düzeyde açıklayıcılar | Kısmi Bağımlılık Profilleri

**Mustafa Cavus, Ph.D.**

 Eskişehir Teknik Üniversitesi - İstatistik Bölümü

 mustafacavus@eskisehir.edu.tr

 [linktr.ee/mustafacavus](https://linktr.ee/mustafacavus)

# Giriş

Kısmi bağımlılık profilleri (*Partial dependence profiles*), Friedman (2000) tarafından *Gradient Boosting Machines* (GBM) yöntemi ile birlikte önerildi.

Ancak son yıllarda bir çok açıklanabilirlik paketinde (**DALEX**, **iml**, **PDP**, **PDPbox**) de yer almasıyla popüler olmaya başladı.

Friedman, Jerome H. 2000. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics* 29: 1189–1232.

# Giriş

Kısmi bağımlılık profillerinin geliştirilmesindeki temel amaç model tahminin beklenen değerinin ilgilenilen bir açıklayıcı değişkenin bir fonksiyonu olarak nasıl davrandığını göstermektir.

Bir model için eğitildiği veri kümesinde yer alan tüm gözlemler veya gözlemlerin herhangi bir alt kümesi için ayrı ayrı oluşturulan *Ceteris-Paribus* profillerinin **birleştirilmesi** ile oluşturulur.

Gözlemlerin alt kümelerine özgü olarak oluşturulan kısmi bağımlılık profilleri model tahminlerinin istikrarı (*stability*) konusunda önemli bilgiler sağlayabilir.

# Giriş

Kısmi bağımlılık profilleri farklı modellerin karşılaştırılmasında 3 önemli bilgi sağlar:

**1. Farklı modeller için profiller arasındaki benzerlik güven vericidir.** Bazı modeller diğerlerinden daha esnektir. Esneklik açısından farklılık gösteren modellerin profilleri benzerse, bunu daha esnek modelin aşırı uyum sağlamadığına ve modellerin aynı ilişkiyi yakaladığına dair bir kanıt olarak değerlendirebiliriz.

# Giriş

**2. Profiller arasındaki farklılık, modeli iyileştirmek için bir yol önerebilir.** Daha basit ya da yorumlanabilir bir modelin profili, esnek bir modelin profiliyle uyuşmuyorsa, bu, yorumlanabilir modeli geliştirmek için kullanılabilecek bir değişken dönüşümü gerekliliğine işaret edebilir. Örneğin, bir rastgele orman modeli, bağımlı değişken ile açıklayıcı değişken arasında doğrusal olmayan bir ilişkiyi gösteriyorsa açıklayıcı değişkenin uygun bir dönüşümü, doğrusal regresyon modelinin uyumunu veya performansını iyileştirebilir.

# Giriş

**3. Model performansının karar sınırlarda değerlendirilmesini sağlar.** Modellerin, bağımlı değişkenin değişim aralığının sınır noktalarında, yani en yüksek veya en düşük değerlerinde farklı davranışlara sahip olabileceği bilinmektedir. Örneğin, rastgele orman modeli tahminlerinin ortalamaya doğru yaklaştığı bilinirken, destek vektör makinelerinin bu noktalarda daha büyük bir varyansa sahip olduğu bilinmektedir. Kısmi bağımlılık profillerinin karşılaştırılması, modellerin sınırlardaki davranışlarındaki farklılıkların anlaşılmasına yardımcı olabilir.

# **Kısmi Bağımlılık Profilleri** ***(Partial Dependence Profiles)***

# Kısmi Bağımlılık Profilleri

Model tahmininin beklenen değerinin, ilgilenilen açıklayıcı değişkenin bir fonksiyonu olarak nasıl davrandığını göstermek için, bir dizi bireysel *ceteris-paribus* profilinin ortalaması kullanılabilir.

Bir CP profilinin, gözlem düzeyindeki bir tahminin açıklayıcı bir değişkene bağımlılığını göstermektedir. Bir PD profili, bir veri kümesindeki tüm gözlemler için CP profillerinin ortalaması alınarak elde edilir.



# Kısmi Bağımlılık Profilleri

Toplamsal (*additive*) modeller için CP profillerinin paralel olduğunu unutmayın: benzer şekle sahiplerdir.

CP profillerinin ortalamasının alınması, profillerin şeklini korurken daha kesin bir tahmin sunar.

Ancak, etkileşim içeren (*non-additive*) modeller için CP profilleri paralel olmayabilir.

# Kısmi Bağımlılık Profilleri

İlgilenilen bir açıklayıcı değişken  $X_j$  ile yanıt değişkeni  $y$  arasındaki ilişkiyi göstermek üzere  $f$  modelinin  $z$  noktasındaki değeri üzerinden bir CP profili hesaplanabilir:

$$g_j(z) = E_{X_{-j}}[f(X^j|z)]$$

Daha sonra herbir gözlem değeri için hesaplanan CP profillerinin ortalamasının alınmasıyla PD profiline ulaşılır:

$$\hat{g}_j(z) = \frac{1}{n} \sum_{i=1}^n f(x_i^j|z)$$

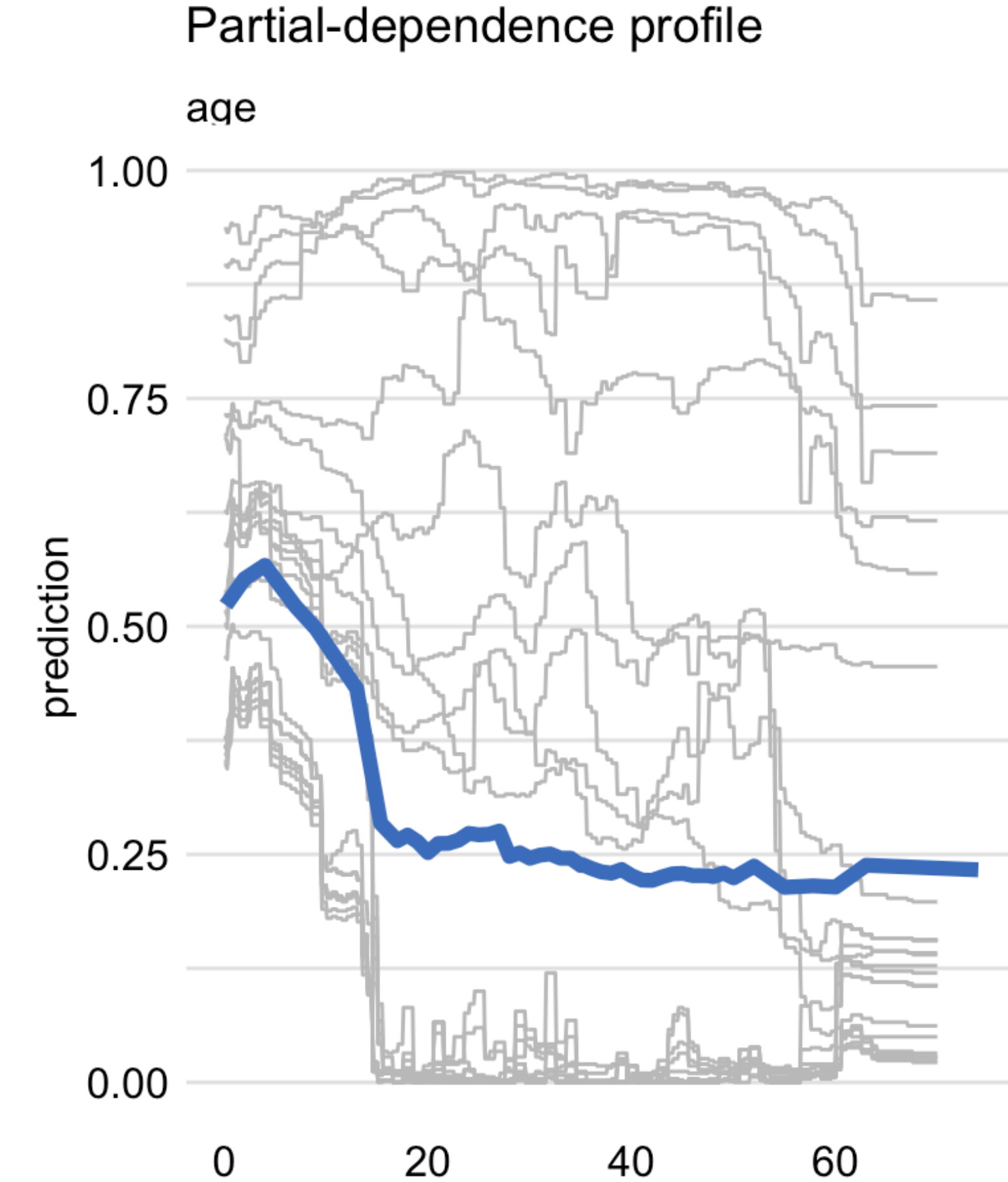
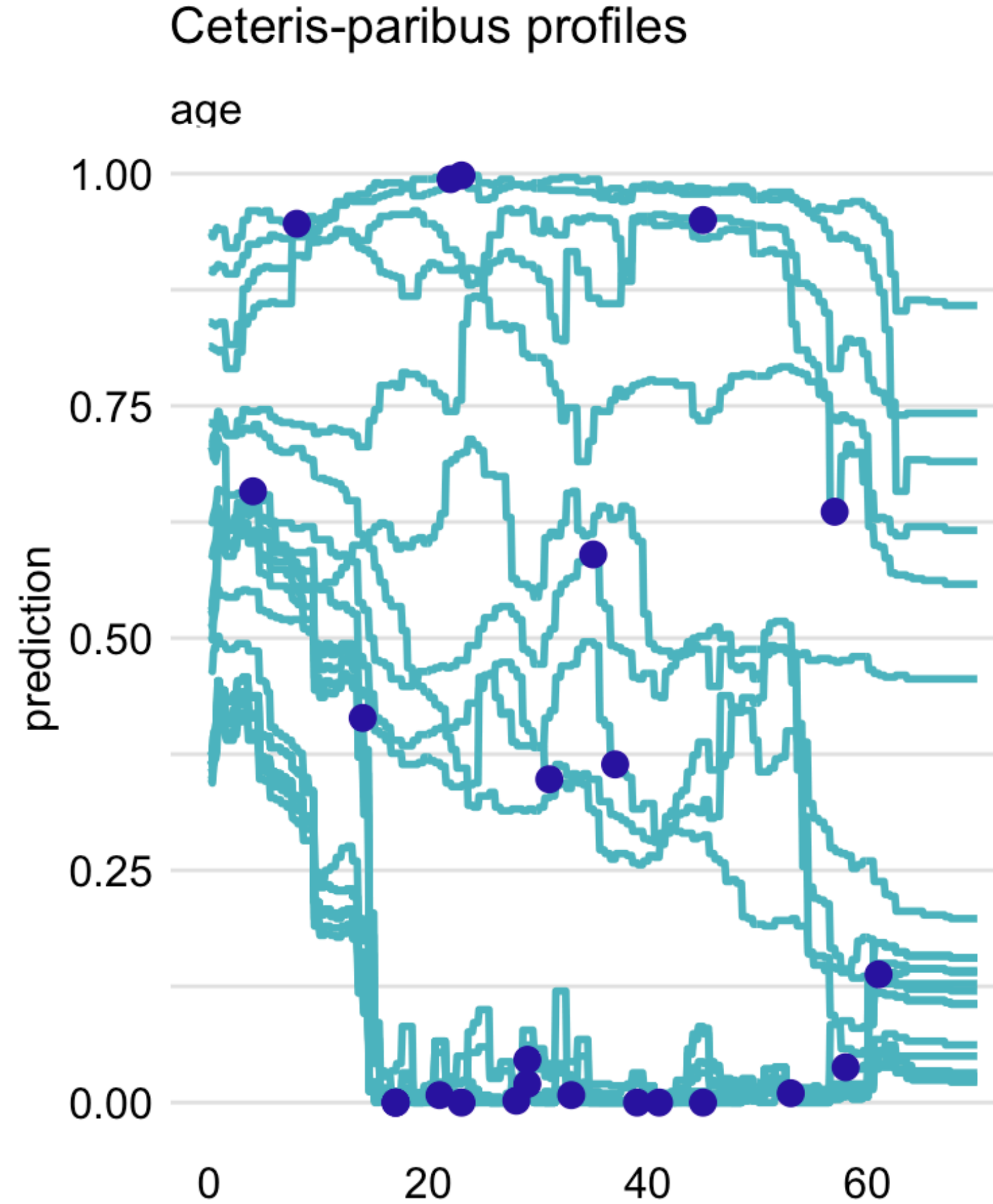
# Örnek

Titanic veri kümesinden rastgele seçilen 25 gözlem için rastgele orman modelindeki açıklayıcı değişken yaş için CP profillerini sunmaktadır.

Açıklayıcı değişkenlerin toplamsal olmayan etkilerini gösteren profillerin paralel olmadığına dikkat edin.

Sağ taraftaki panel, PD profilinin bir tahminini sunan CP profillerinin ortalamasını gösterir. Açıkçası, PD profilinin şekli, grafiğin üst kısmında gösterilen beş CP profilinden oluşan grubun şeklini yansıtmamaktadır.

Bununla birlikte, CP profillerinin çoğunluğunun 2 ile 18 yaş arasındakiler için tahmin edilen hayatta kalma olasılığında önemli bir düşüşe işaret ettiği gerçeğini yansıtıyor gibi görünüyor.



# Kısmi Bağımlılık Profili Türleri

# Kümelenmiş Kısmı Bağımlılık Profilleri

Daha önce de belirtildiği gibi, profiller paralelse CP profillerinin ortalaması iyi bir özetir.

Paralel değilse ortalama, profillerin bir alt kümesinin şeklini yeterince temsil etmeyebilir.

Bu sorunla başa çıkmak için profillerin kümelenmesi ve ortalamanın her küme için ayrı ayrı hesaplanması düşünülebilir.

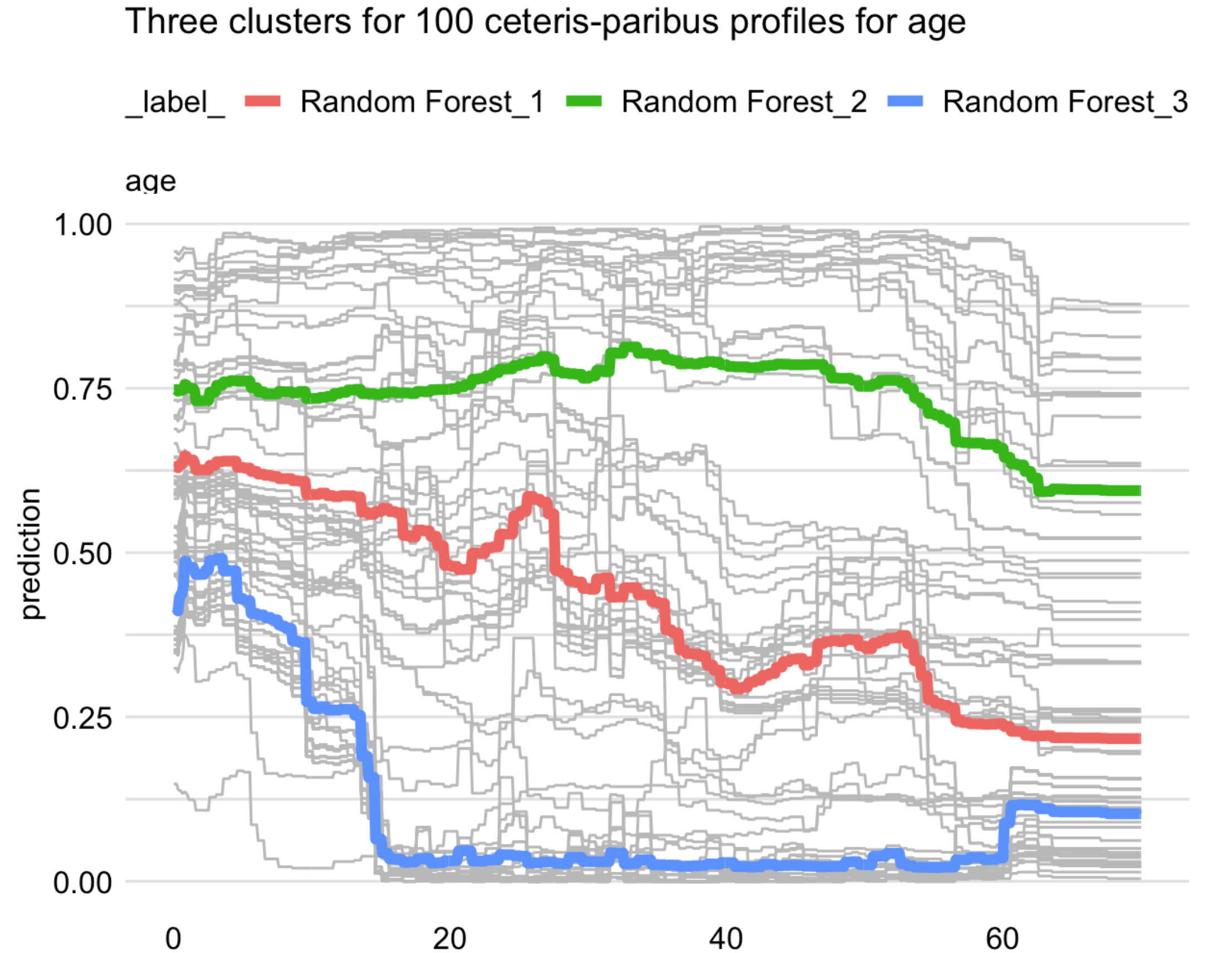
CP profillerini kümelemek için K-ortalamları veya hiyerarşik kümeleme gibi standart yöntemler kullanılabilir. Gözlemler arasındaki benzerlikler CP profilleri arasındaki Öklid mesafesine göre hesaplanabilir.

# Kümelenmiş Kısmı Bağımlılık Profilleri

Titanic veri setinden rastgele seçilen 100 gözlem için bu yaklaşımın rastgele orman modeline uygulanmasını göstermektedir.

**Age** değişkenine ilişkin CP profilleri gri renkle gösterilmiştir. Bunların üç kümeye ayrılabilir:

1. Birincisi, 18 yaşın üstündeki yolcular için tahmin edilen hayatta kalma olasılığında önemli bir düşüş olan bir grup yolcu (**mavi**)
2. Diğeri ise yaş ve ortalama hayatta kalma olasılığının neredeyse doğrusal bir azalma gözleendiği grubu (**kırmızı**)
3. Neredeyse yaşa göre ortalama hayatta kalma olasılığının sabit olduğu grup (**yeşil**).





# Gruplanmış Kısmı Bağımlılık Profilleri

İlgilenilen açıklayıcı değişken için CP profillerinin şeklini etkileyen bir başka açıklayıcı değişken olabilir.

En belirgin durum, modelin bir değişken ile bir diğeri arasında etkileşim olmasıdır.

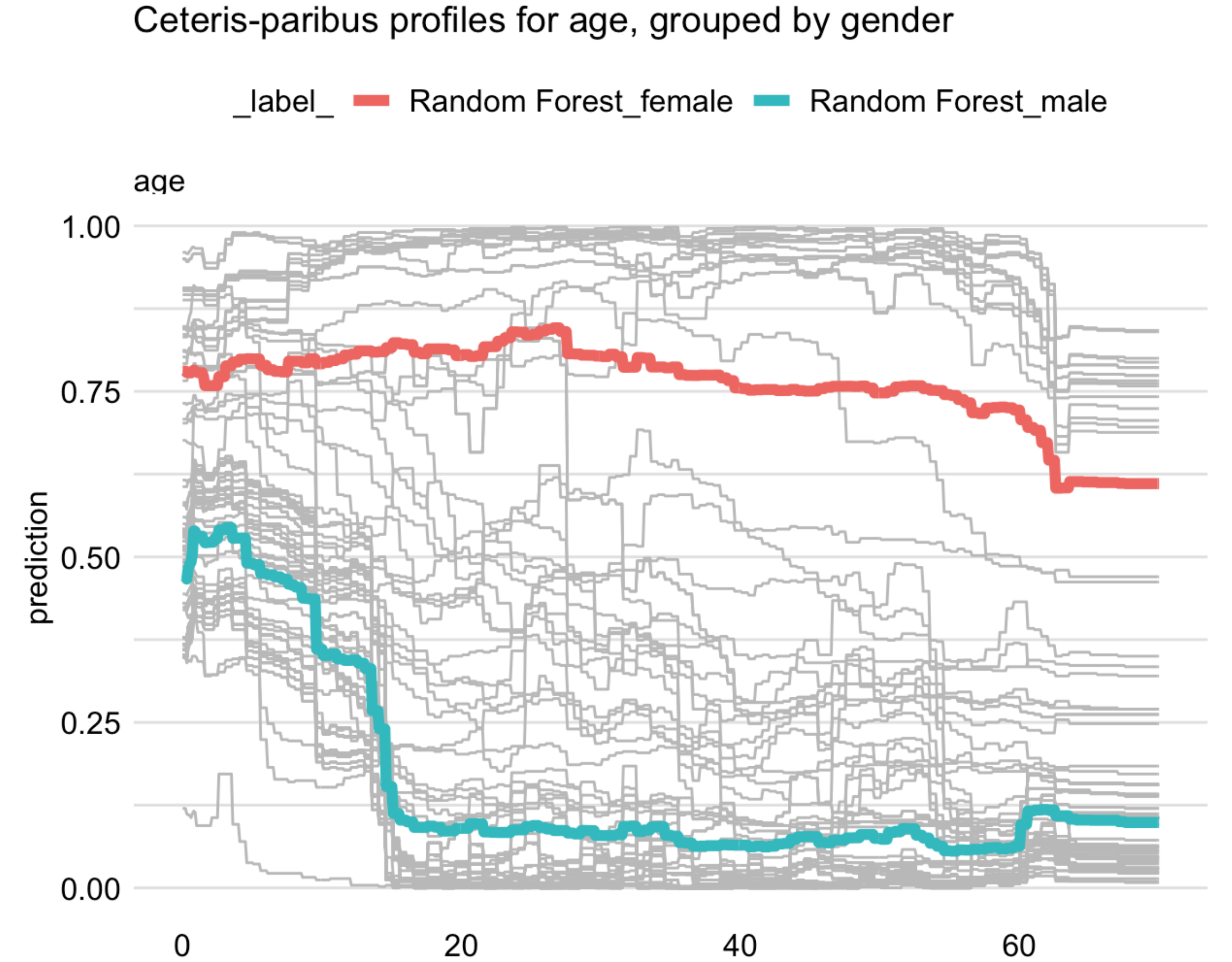
Bu durumda, etkileşimde yer alan değişken tarafından tanımlanan gözlem grupları içerisinde ilgilenilen değişkene ilişkin PD profillerinin incelenmesi gereklidir.

# Gruplanmış Kısmı Bağımlılık Profilleri

Titanic veri kümesinden rastgele seçilen 100 gözlem için rastgele orman modeli uygulanmasında açıklayıcı değişken **age** için CP profilleri gri renkle verilmiştir.

Kırmızı ve mavi renkle sırasıyla kadınlar ve erkekler için PD profilleri görülmektedir.

Cinsiyete özel ortalamalar farklı şekillerdedir: **Kadınlar için tahmin edilen hayatta kalma olasılığı, erkeklerle karşılaştırıldığında farklı yaşlarda daha istikrarlıdır.** Dolayısıyla PD profilleri yaş ve cinsiyet arasındaki etkileşimi açıkça göstermektedir.





# Karşılaştırmalı Kısmı Bağımlılık Profilleri

Tek bir model için kümelenmiş veya gruplandırılmış PD profillerinin karşılaştırılması, modelin tahminlerinin istikrarı hakkında önemli bilgiler sağlayabilir.

Aynı veri kümesi üzerinde eğitilmiş farklı modellerin PD profillerinin karşılaştırılması da model farklılıklarına ilişkin önemli bilgiler sağlayabilir.

# Karşılaştırmalı Kısmi Bağımlılık Profilleri

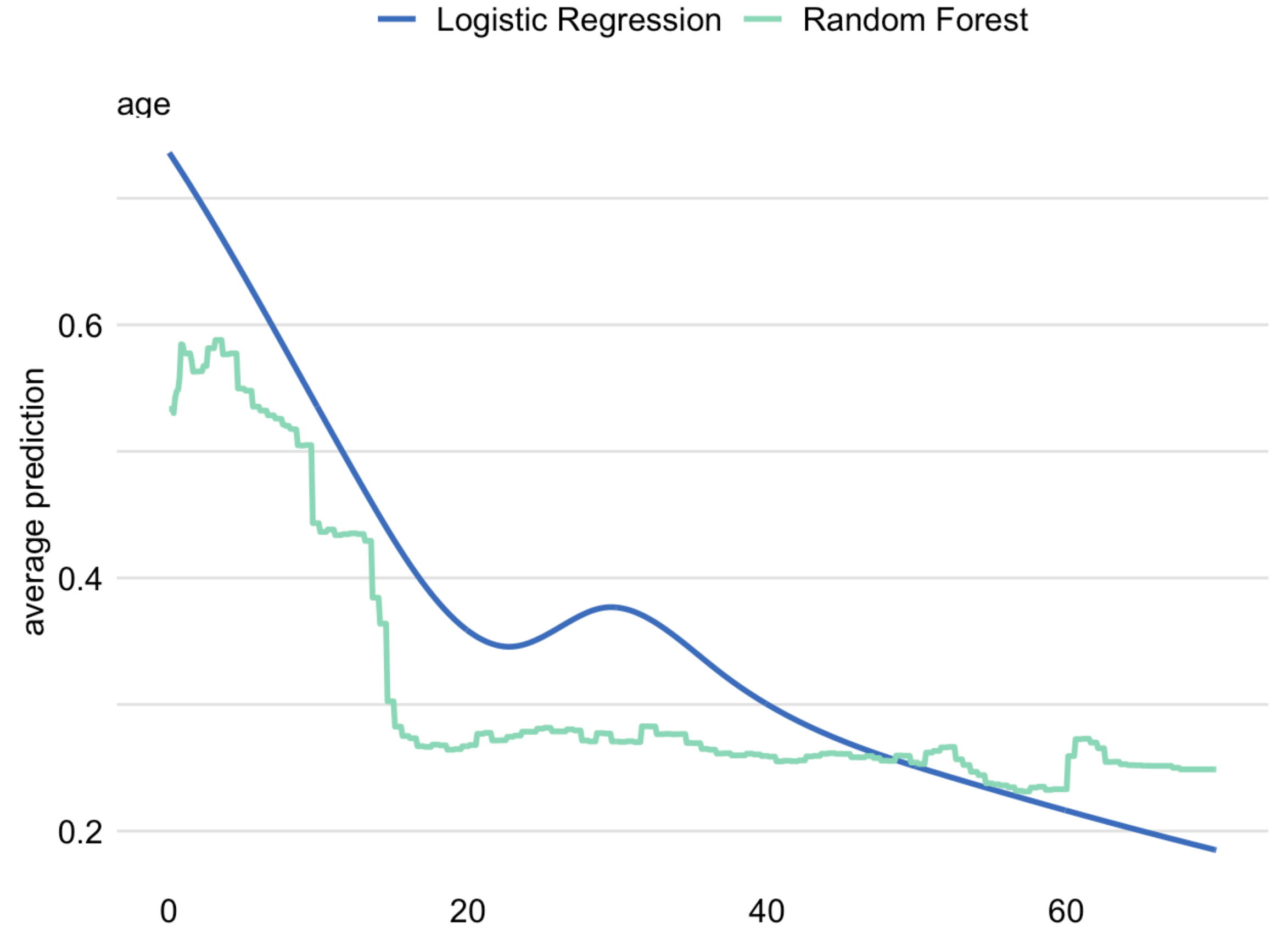
Titanic verileri üzerinde eğitilen bir lojistik regresyon modeli ve rastgele orman modeli için **age** değişkenine karşılık oluşturulan PD profillerini grafikte verilmiştir.

Profiller, yaş ile tahmin edilen hayatta kalma olasılığı arasındaki genel ilişki açısından benzerdir (yolcu ne kadar gençse, hayatta kalma şansı da o kadar yüksek olur).

Ancak rastgele orman modelinin profili daha pürüzlüdür. Her iki model arasındaki fark, yaş değişkeninin en düşük değerleri aldığı bölgede en büyüktür.

Bu beklenen bir durumdur çünkü rastgele orman modelleri genel olarak tahminleri ortalamaya yaklaştırır ve eğitim veri setinde gözlemlenen değer aralığı dışındaki gözlemler üzerinde gerçekleştirilecek tahminler için çok kullanışlı değildir.

Partial-dependence profiles for age for two models



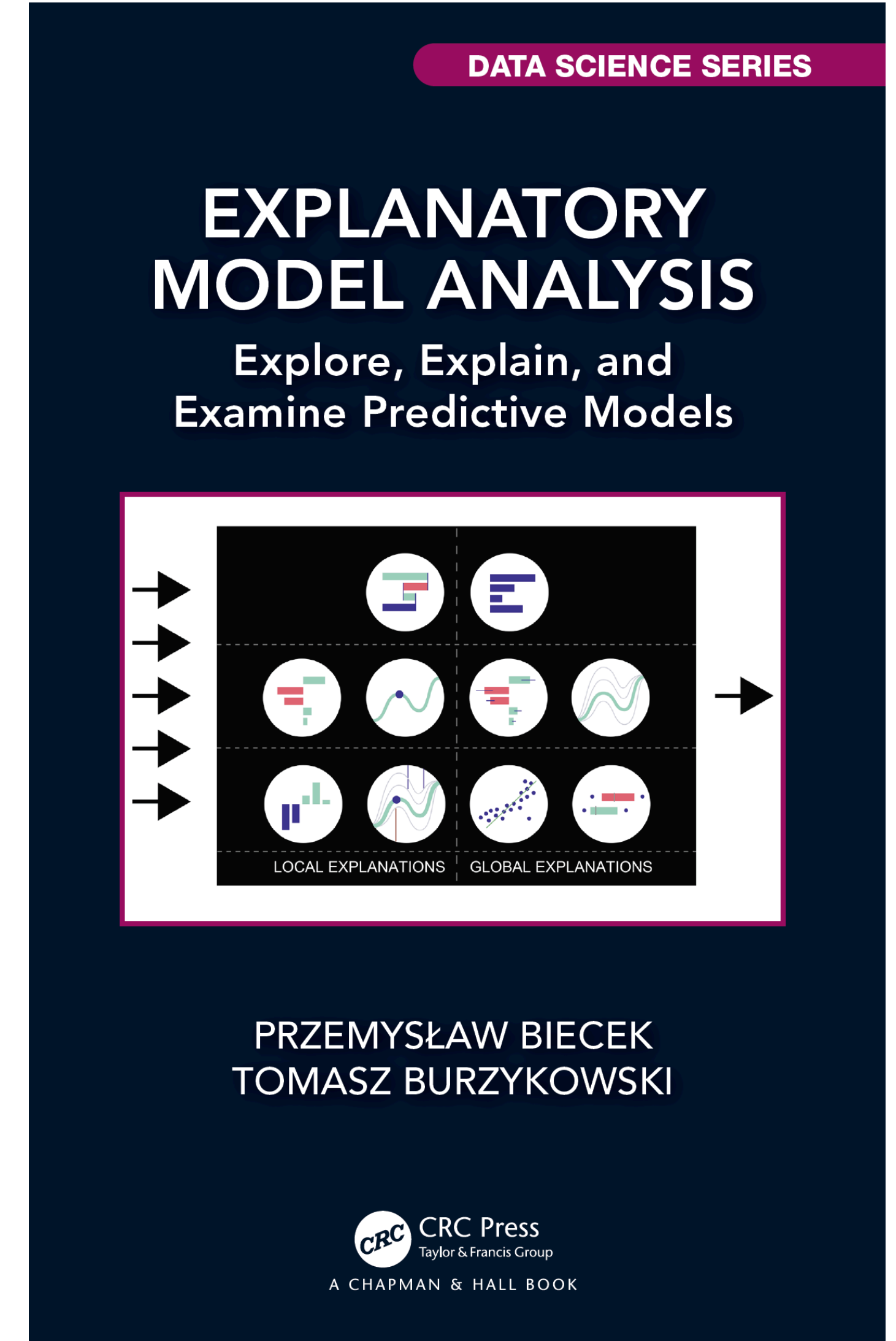
# Artı ve eksileri

- Modelden bağımsız
- Görselleştirilebilir
- Anlaşılması kolay
- Gözlemlerin alt kümeleri için de oluşturulabilirler.
- Farklı modellerin davranışını karşılaştırmak için kullanılabilirler.

Toplamsal olmayan modellerde yer alan değişkenler ve yanıt değişkeni arasındaki ilişki konusunda yanıtıcı bilgiler verebilir.

# Kaynaklar

Ders materyallerinin hazırlanmasında **Explanatory Model Analysis (Biecek and Burzykowski, 2021)** kitabından yararlanılmıştır. Kitabın ücretsiz online versiyonuna bağlantı üzerinden erişilebilir: <https://ema.drwhy.ai/>



Ders notlarına dersin **GitHub** reposu üzerinden ulaşabilirsiniz.

Ders ile ilgili sorularınız için **mustafacavus@eskisehir.edu.tr** adresi üzerinden benimle iletişime geçebilirsiniz.

**Mustafa Cavus, Ph.D.**

 Eskişehir Teknik Üniversitesi - İstatistik Bölümü

 mustafacavus@eskisehir.edu.tr

 linktr.ee/mustafacavus