4 Nisan 2023

Açıklanabilir Yapay Zeka

5. Hafta: Lokal düzeyde açıklayıcılar | Ceteris-paribus yöntemi

Mustafa Cavus, Ph.D.
🏢 Eskişehir Teknik Üniversitesi - İstatistik Bölümü
📧 mustafacavus@eskisehir.edu.tr
🌐 linktr.ee/mustafacavus

Giriş

- Bugüne kadar ele alınan açıklayıcılar, bir model tahmine modelde yer alan değişkenlerin katkısını açıklamak üzerine kurgulanmışlardır.
- Bu derste ise ilgilenilen bir değişkenin model tahminine olan katkısını incelemeye yarayan **Ceteris-paribus** yöntemine odaklanacağız.

Ceteris-paribus ilkesi

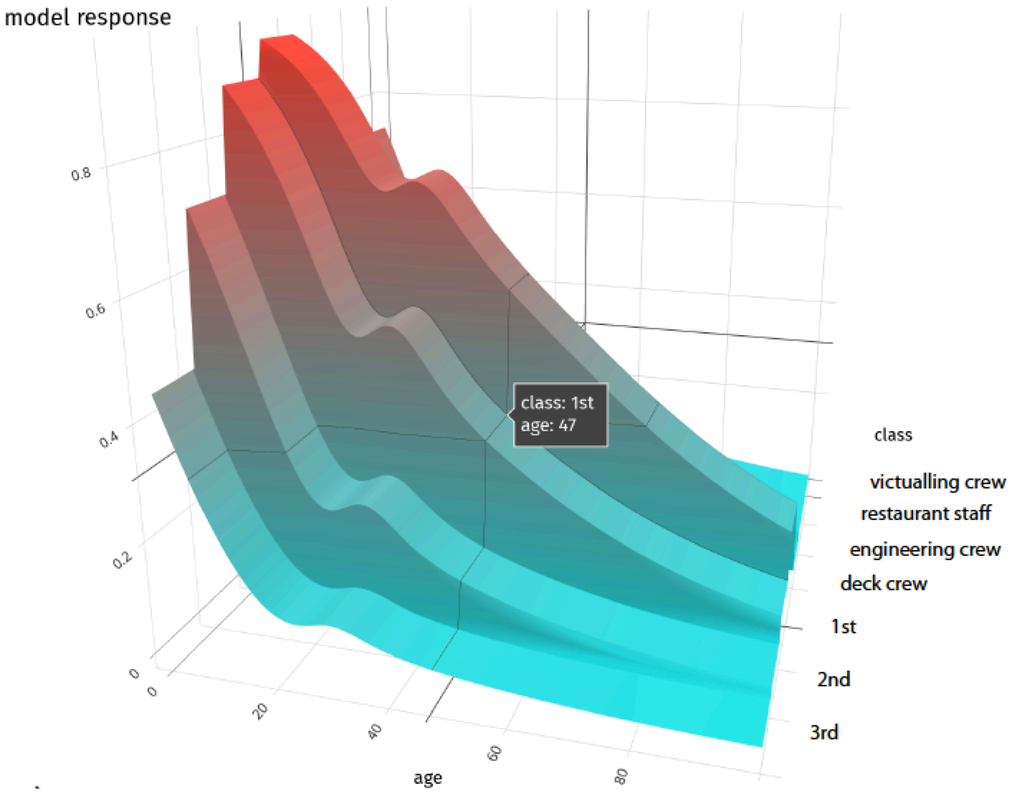
- *Ceteris-paribus*, ekonomide kullanılan ve modelleme yaklaşımına devşirilen bir terimdir.
- Ekonomide, diğer tüm koşullar sabitken ilgilenilen durumdaki değişimleri inceler.
- Kabaca *ceteris-paribus* ilkesi, **diğer tüm koşullar sabitken** anlamına gelir.

Ceteris-paribus yöntemi

- *Ceteris-paribus* profilleri, ilgilenilen değişkenin aldığı değer değiştiğinde bir modelin tahminin nasıl değişeceğini gösterir.
- Matematiksel olarak bağımlı değişkenin koşullu beklenen değerinin, belirli bir açıklayıcı değişkenin değerlerine bağımlılığını gösterir.

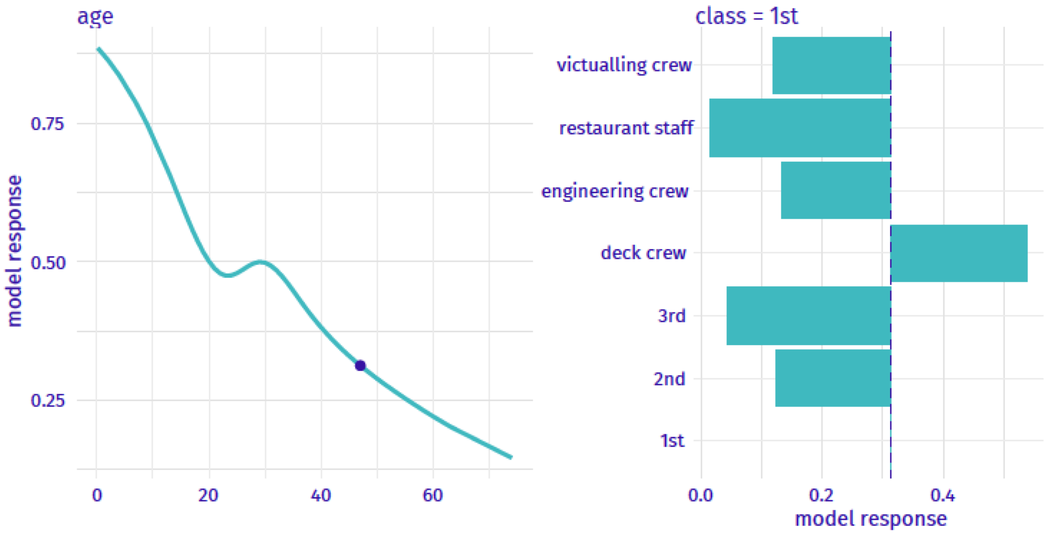
Örnek

Yandaki grafikte, Titanic veri seti üzerinde eğitilen bir lojistik regresyon modeli için age ve class değişkenlerinin değerlerine karşılık model tahminin değişimini gösteren *ceteris-paribus* yüzeyleri verilmiştir.



Örnek

Yandaki grafikte ise age ve class değişkenleri için *ceteris-paribus* profilleri verilmiştir.



Ceteris-paribus yöntemi

i gözlemi için açıklayıcı değişkenlerin aldığı değerlerin vektörü $\underline{x}_i$, rasgele gözlem değerlerinden oluşan vektör $\underline{x}_*$ ve ilgilenilen j değişkeninin değerlerinin yer aldığı sütun $\underline{x}^j_*$ ile gösterilmek üzere, $\underline{x}^{-j}_*$ j değişkeninin çıkarıldığı durumu temsil eder. Bu durumda $\underline{x}^{-j|=z}_*$ ise çıkarılan j değişkenin değerlerinin yerine herhangi değerlerin yerleştirildiği durumu temsil eder.

$h()$ ile gösterilen $f()$ modeli için tek boyutlu bir ceteris-paribus profili, ilgilenilen j değişkeni için bir $\underline{x}_*$ gözlemi için matematiksel olarak aşağıdaki gibi gösterilir:

$$h^{fij}_{\underline{x}_*}(z) = f(\underline{x}^{j|=z}_*)$$

Ceteris-paribus yöntemi

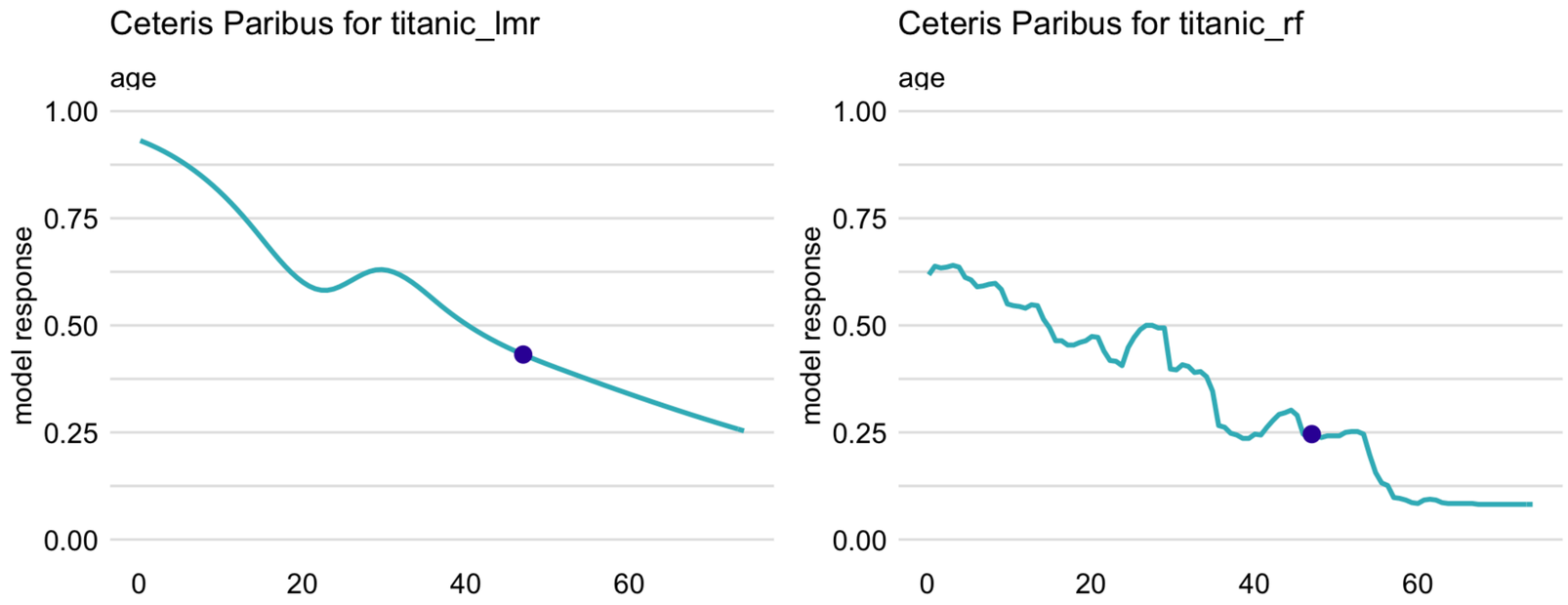
$$h^{fij}_{\underline{x}_*}(z) = f(\underline{x}^{j|=z}_*)$$

burada görüldüğü üzere ceteris-paribus profili h, bir değişken için modelin koşullu beklenen değerinin bağımlılığını gösteren bir fonksiyondur.

Fonksiyonun değeri hesaplanırken ilgilenin j değişkeni dışında kalan tüm değişkenlerin değerinin sabit ve gözlendiği değerlere eşit olduğunun altını çizmek, açıklayıcıyı kullanırken sınırlarının bilinmesi açısından önemlidir.

Örnek

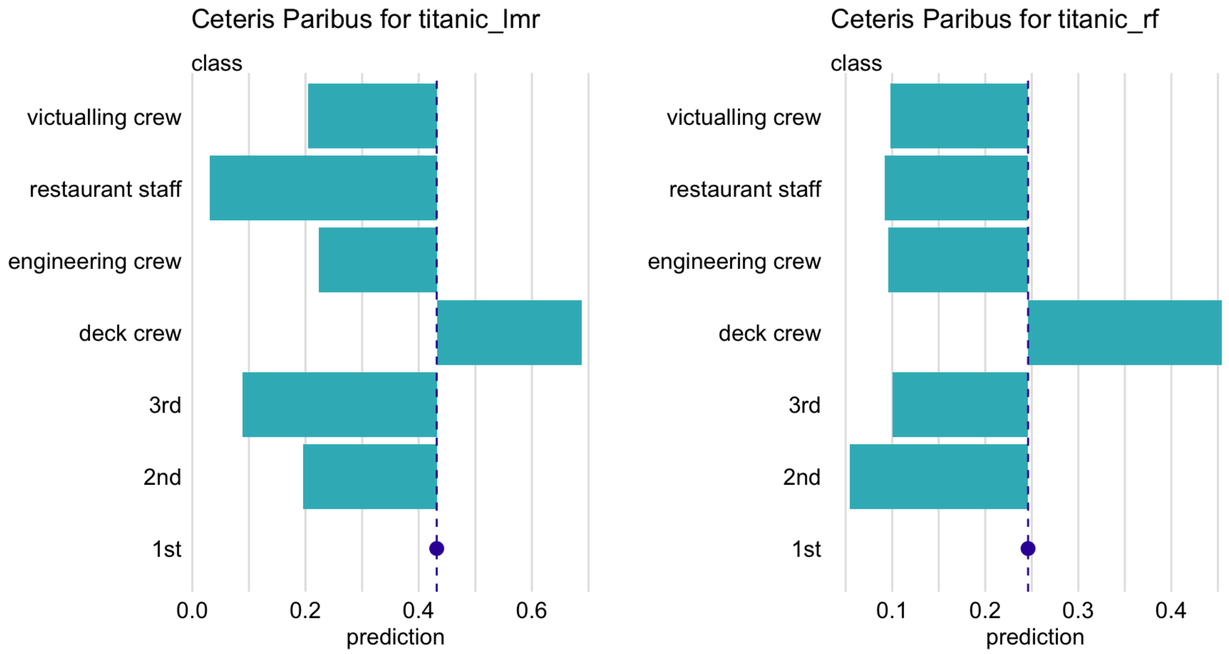
Aşağıdaki grafikte, Titanic veri setinde yer alan yolcu Henry için lojistik regresyon (solda) ve rasgele orman (sağda) modelleri üzerinden *age* değişkeni için oluşturulan *ceteris-paribus* profilleri verilmiştir.



Genel olarak lojistik regresyon modeline age değişkenin değerlerindeki değişim doğrusala yakın bir görüntüde katkı sunarken, rasgele orman modelinde ise katkının doğrusallıktan uzak ve değişkenlik gösteren bir yapıda olduğu görülmektedir.

Örnek

Aşağıdaki grafikte, Titanic veri setinde yer alan yolcu Henry için lojistik regresyon (solda) ve rasgele orman (sağda) modelleri üzerinden kategorik yapıda olan *class* değişkeni için oluşturulan *ceteris-paribus* profilleri verilmiştir.



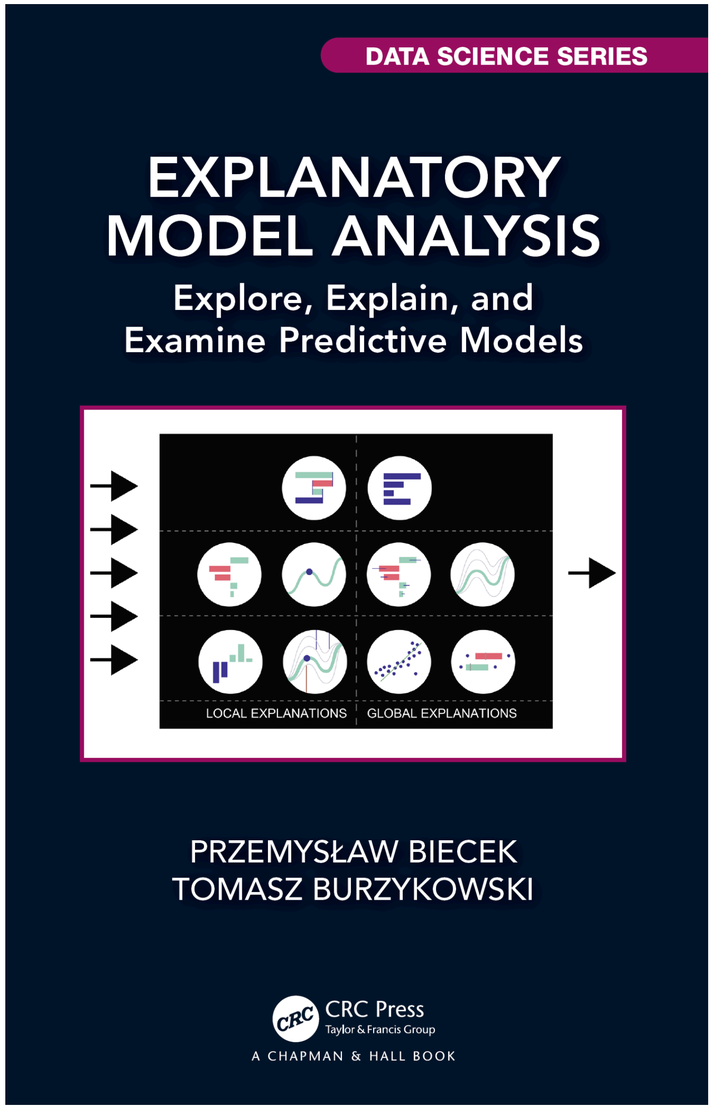
Artı ve eksileri

- Grafiksel bir gösterim sunduğu için anlaşılması ve açıklanması kolaydır.
- Birden fazla model için tek bir grafik üzerinde karşılaştırma yapabilmek mümkün olduğu için kullanışlıdır.
- Model tahminin kararlılığını (değişmezlik) anlamak için faydalıdır.

- Değişkenler arasında ilişki olmazı durumunda yanıltıcı sonuçlar verebilir.
- Çok sayıda değişken içeren modeller söz konusu olduğunda kullanışlı olmazlar.
- Kategorik değişkenlerin düzey sayısı arttığında da benzer bir kullanışsızlık durumu ortaya çıkabilir.

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** sayfası 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