

18 Nisan 2023

Açıklanabilir Yapay Zeka

6. Hafta: Lokal teşhis grafikleri

Mustafa Cavus, Ph.D.

 Eskişehir Teknik Üniversitesi - İstatistik Bölümü

 mustafacavus@eskisehir.edu.tr

 linktr.ee/mustafacavus

Giriş

- Bir modelin tahmin performansı genel olarak tatmin edici düzeyde olmasına rağmen, bazı gözlemler için çok düşük düzeyde kalabilir.
- Örneğin, bir hastanede yatan hastalar için eğitilen bir model, başka bir hastanedeki hastalar için iyi bir performans göstermeyebilir. Ya da yaz tatili için tüketici kredi riskini tahmin etmek için geliştirilen bir model, kış tatili için tüketici kredini riskini tahmin etmekte başarılı olmayabilir.
- Bu nedenle, kritik kararlar söz konusu olduğunda, ilgili gözlemlere benzer gözlemler için modelin lokal olarak nasıl davrandığını kontrol etmek önemlidir.

Giriş

Bu derste, bu soruna odaklanan iki lokal teşhis grafiğini ele alacağız:

1. Yerel doğruluk (fidelity) grafiği
2. Yerel kararlılık (stability) grafiği

Komşuluğun belirlenmesi

- Her iki yöntemin de komşuluklar üzerinde çalışması nedeniyle, öncelikle bu konuya odaklanmak gerekir.
- En yakın komşuluğu belirlemeden önce iki önemli sorunun yanıtlanması gerekmektedir:
 1. Komşuluk kaç gözlemden oluşmalı?
 2. Gözlemlerin komşuluğunu ölçmek için hangi metrik kullanılmalı?

Komşuluğun belirlenmesi

Komşuluk sayısının belirlenmesi:

- Komşuların sayısının düşük tutulması daha lokal bir analiz yapılmasını demektir.
- Ancak az sayıda komşuluğun seçimi yüksek varyansa neden olabilir.
- Genel olarak komşuluk sayısının 20 olması kabul edilebilir bir düzeydir.

Komşuluğun belirlenmesi

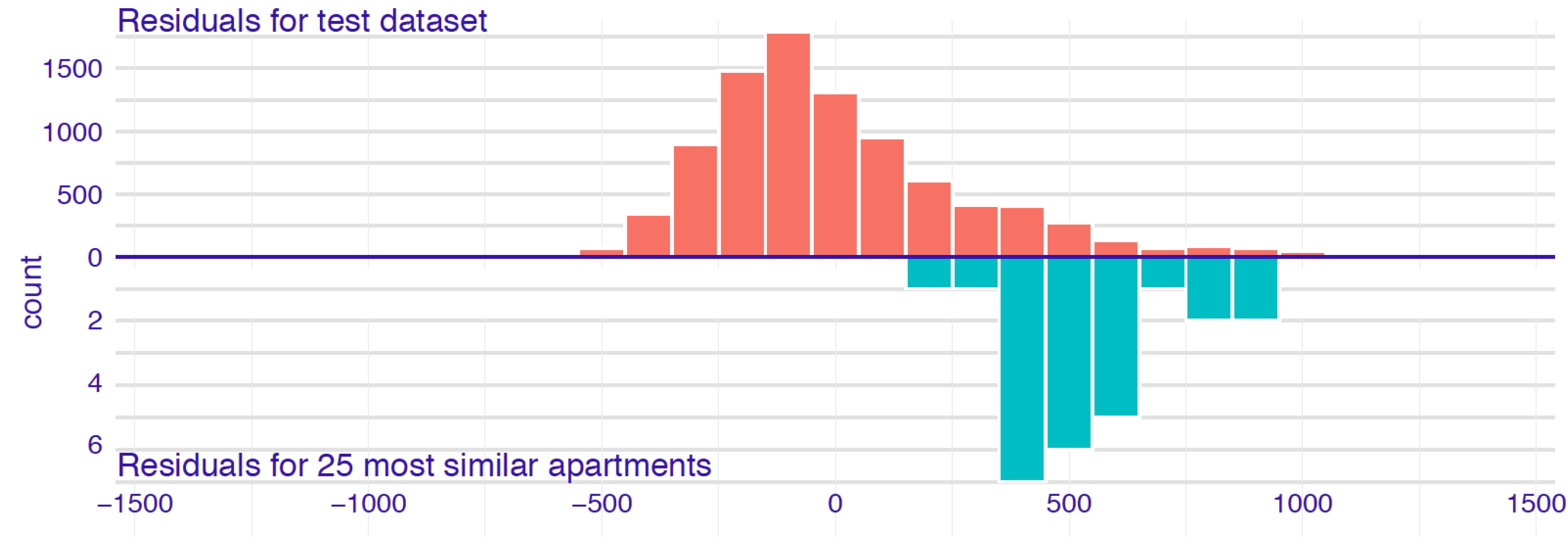
Komşuluk ölçütünün belirlenmesi:

- Komşuluğun belirlenmesinde kullanılan benzerlik ölçüsünün seçimi önemlidir. Ölçütün değişken önemliliklerini dikkate alması ve farklı veri tipleri üzerinde çalışabilmesi önemlidir.
- Bu iki yöntemde Gower benzerlik ölçüsü kullanılmıştır.
- Gower ölçütünün avantajı, hem kategorik hem de sürekli değişkenler üzerinde çalışabilmesidir.
- Dezavantajı ise değişkenler arasındaki korelasyonu ve değişkenlerin önemliliğini dikkate almamasıdır.

1. Yerel doğruluk grafiği

Grafikte tüm gözlemler ve ilgilenilen gözlemin komşuluğunda yer alan gözlemlerin artıklarının dağılımları karşılaştırılmaktadır.

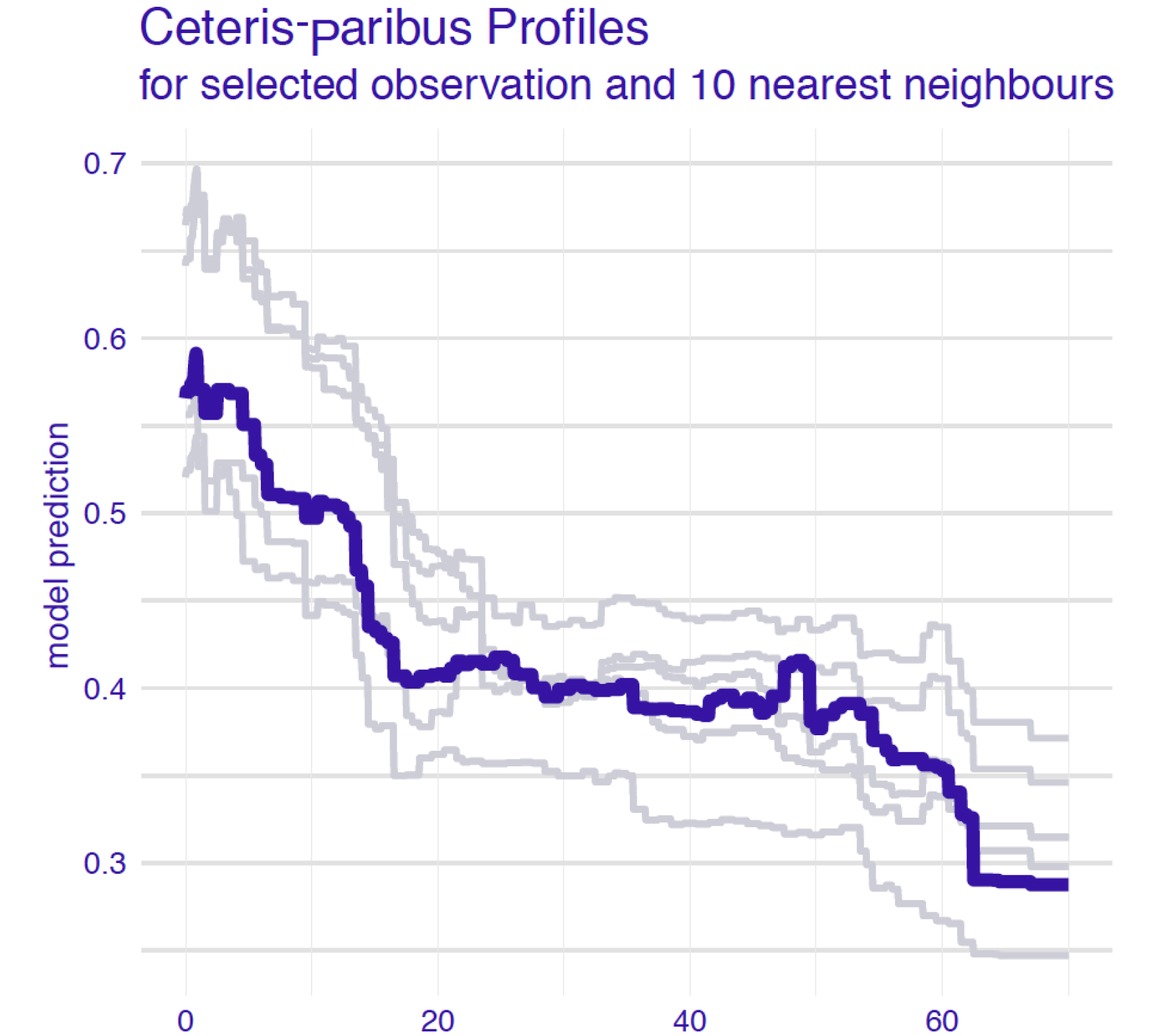
- Normal şartlar altında, iki grubun dağılımının benzer olması beklenmektedir.
- İkinci grup dağılımının negatif ya da pozitif yönde kaymış olması modelde yerel problemler olduğuna bir işarettir.



Görsel inceleme dışında, Wilcoxon testi gibi istatistiksel yöntemlerden de faydalanılabilir. Ancak karşılaştırılan iki grubun bağımsız olmasına dikkat edilmelidir.

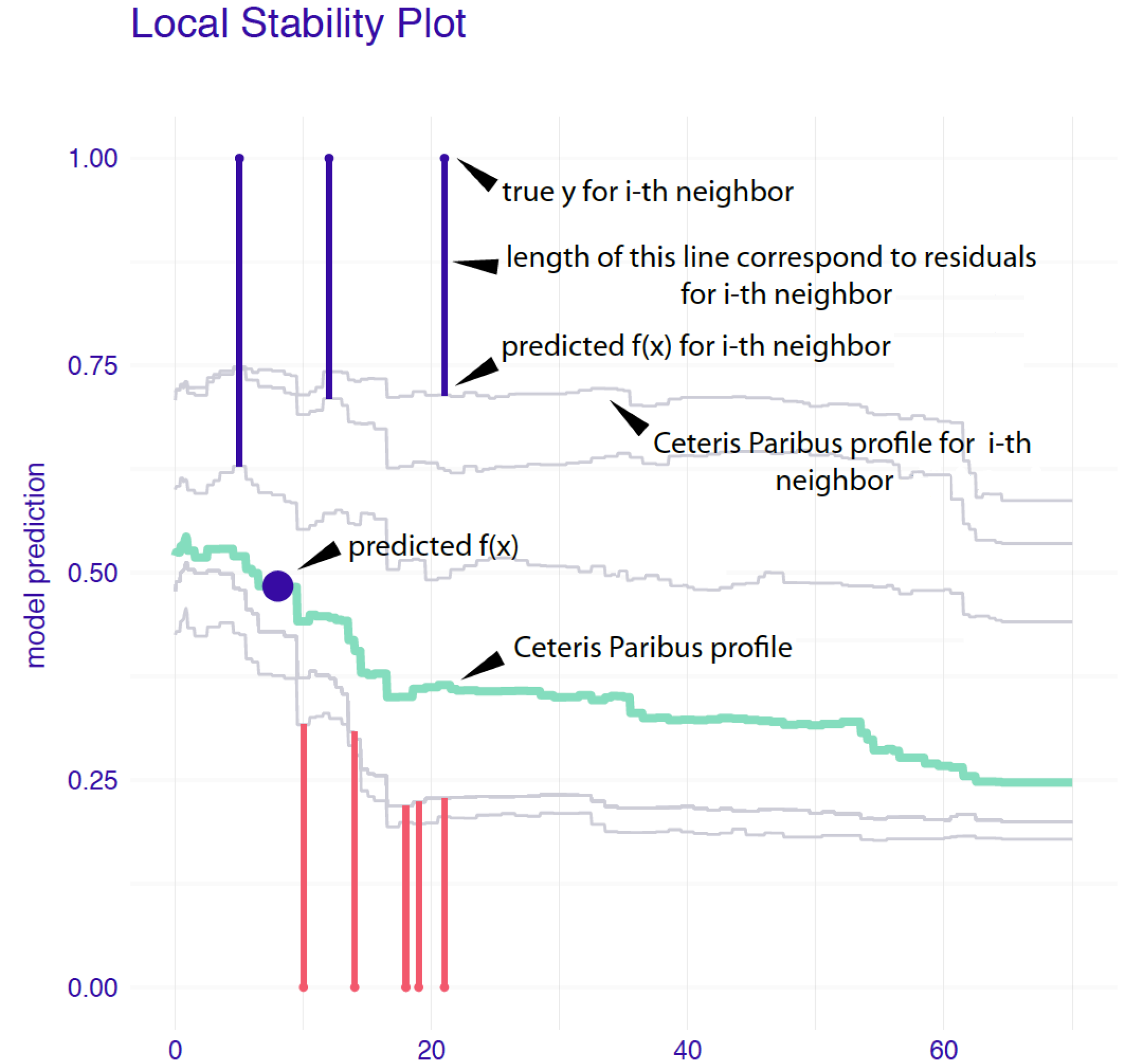
2. Yerel kararlılık grafiği

- İlgilenilen gözlemin komşuluğunda yer alan gözlemler belirlendikten sonra, her bir gözlem ve istenilen açıklayıcı değişkenler için CP profilleri grafiksel olarak karşılaştırılabilir.
- Çok sayıda değişken içeren modeller için, değişken önemliliği ölçüsünü kullanarak en önemli belirli sayıda değişken için bu strateji uygulanabilir.



2. Yerel kararlılık grafiği

- CP profilleri model kararlılığını değerlendirmek için kullanışlıdır. Ek olarak, artıkların da eklenmesiyle modelin yerel kararlılığını incelenmesi geliştirilebilir.
- CP profillerine karşılık gelen artıkların eklenmesi ile bir modele ilişkin yerel kararlılık grafiği elde edilebilir.



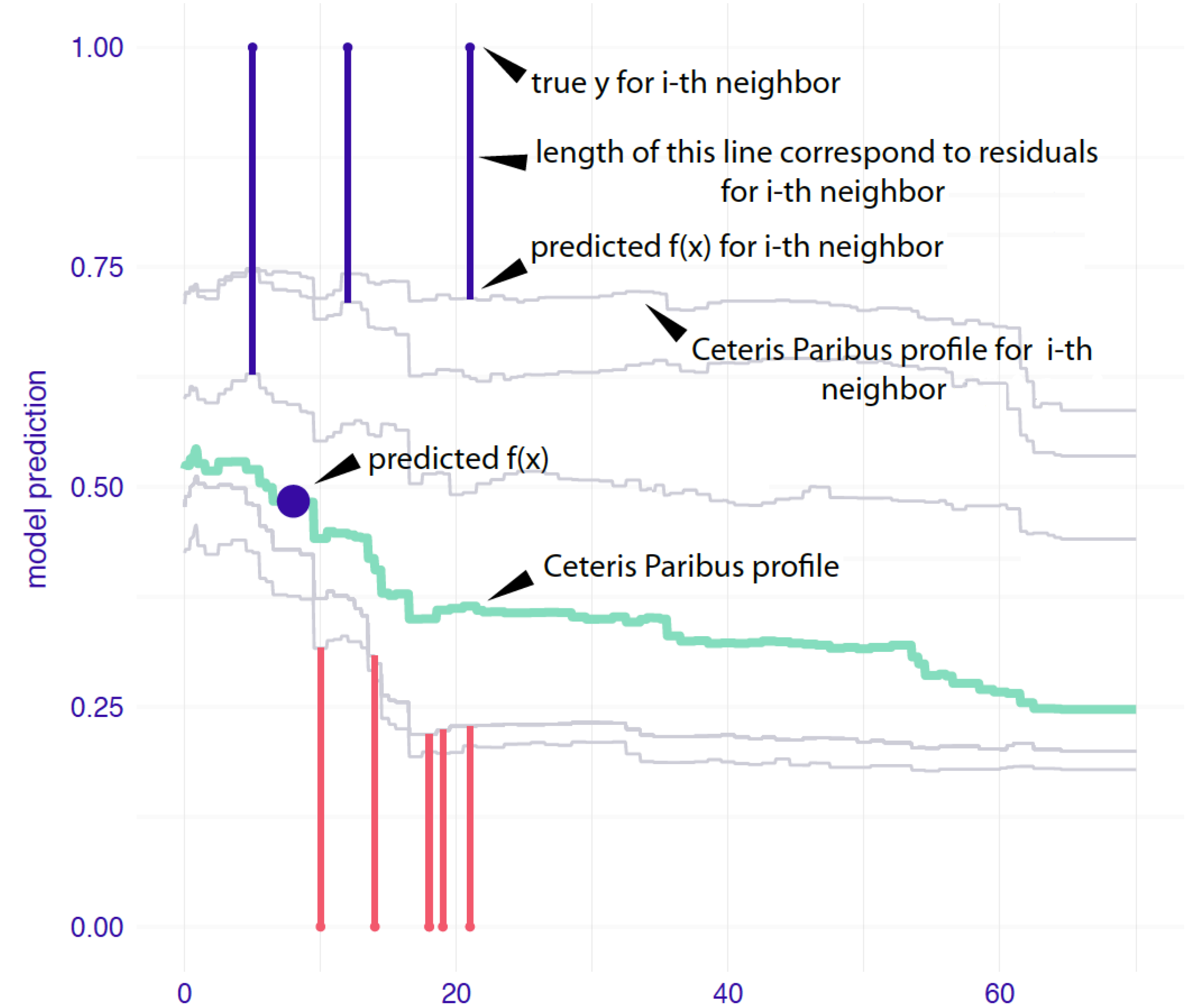
Örnek-1

Titanic veri seti üzerinde eğitilen bir rasgele orman modelini inceleyelim. Bunun için Johnny D gözlemini ele alalım ve ***age*** değişkeni özelinde lokal teşhise devam edelim.

Örnek-1

- Mavi çubuklar pozitif artıkları, kırmızı çubuklar ise negatif artıkları göstermektedir.
- Artıkları temsil eden çubukların daha kısa olduğu noktalarda modelin daha doğru tahminler ürettiğini söylenebilir.
- Modelin kararlı olduğunu söyleyebilmek için CP profillerinin birbirine yakın olması gerekmektedir.
- Grafikte görüldüğü üzere model tahminlerinin kararlı olmadığı ortadadır.
- Hem pozitif hem de negatif artıkların gözlenmesi modelin yanlış tahminler üretmediğine bir işarettir.

Local Stability Plot

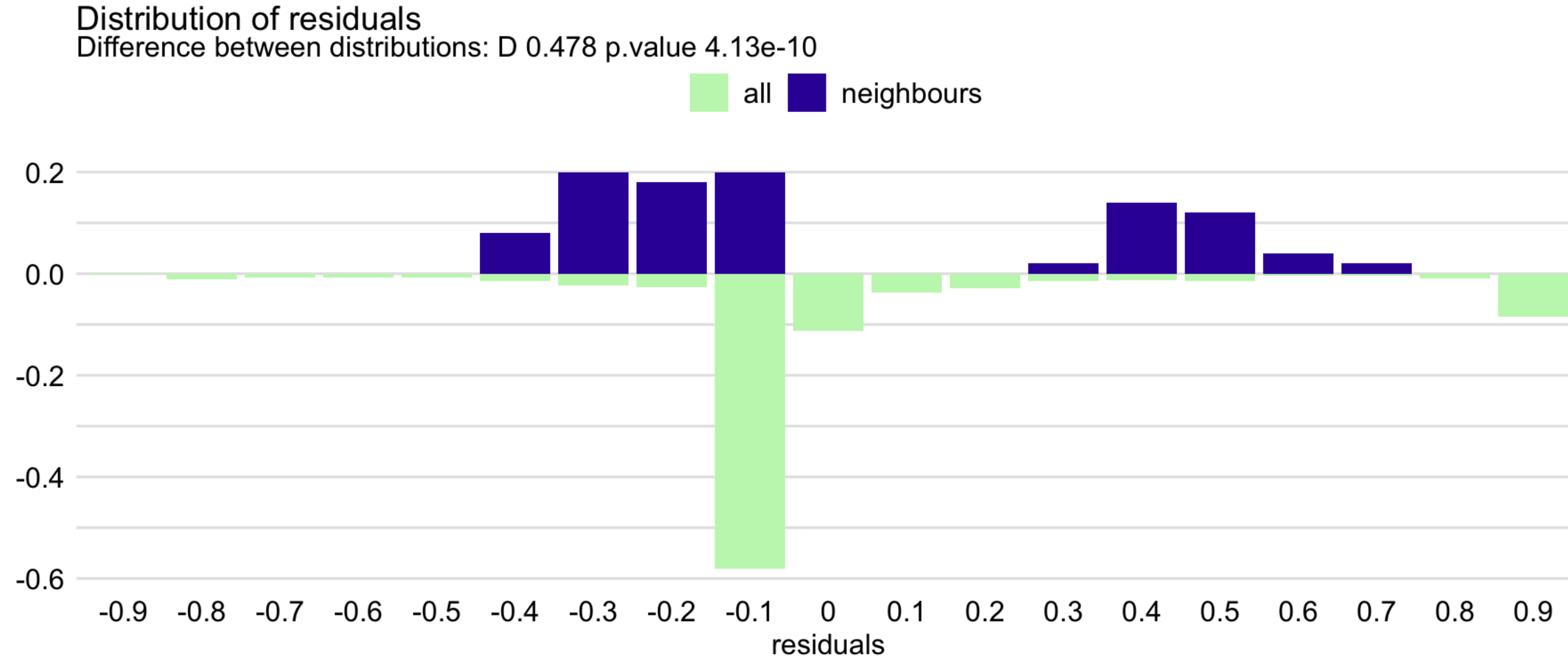


Örnek-2

Aynı modelin yerel performansını, yolcu Henry için 100 gözlem komşuluğunda inceleyelim.

Örnek-2

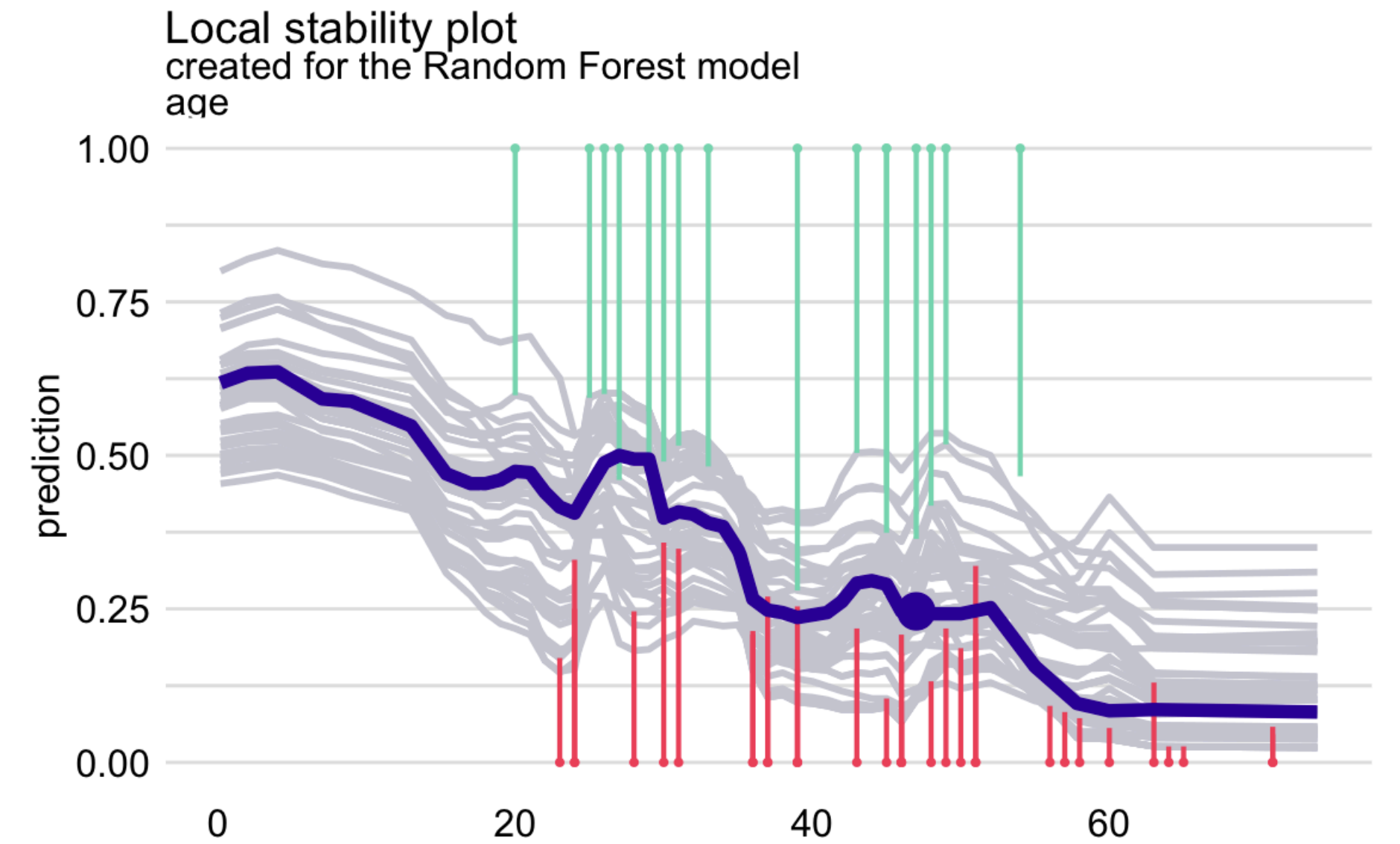
Aynı modelin yerel performansını, yolcu Henry için 100 gözlem komşuluğunda inceleyelim.



Henry için modelin geri kalan gözlemlere göre istatistiksel olarak anlamlı bir farklılıkla artık değerleri oluşturduğu görülmektedir.

Örnek-2

- Aynı gözlem için modelin yerel kararlılığı incelendiğinde, profiller birbirine çok yakın olduğu için tahminlerin kararlı olduğu söylenebilir.
- Negatif artıkların, pozitif artıklarından daha az olması modelin pozitif yönde bir yanlılığa sahip olduğuna işaret olabilir.



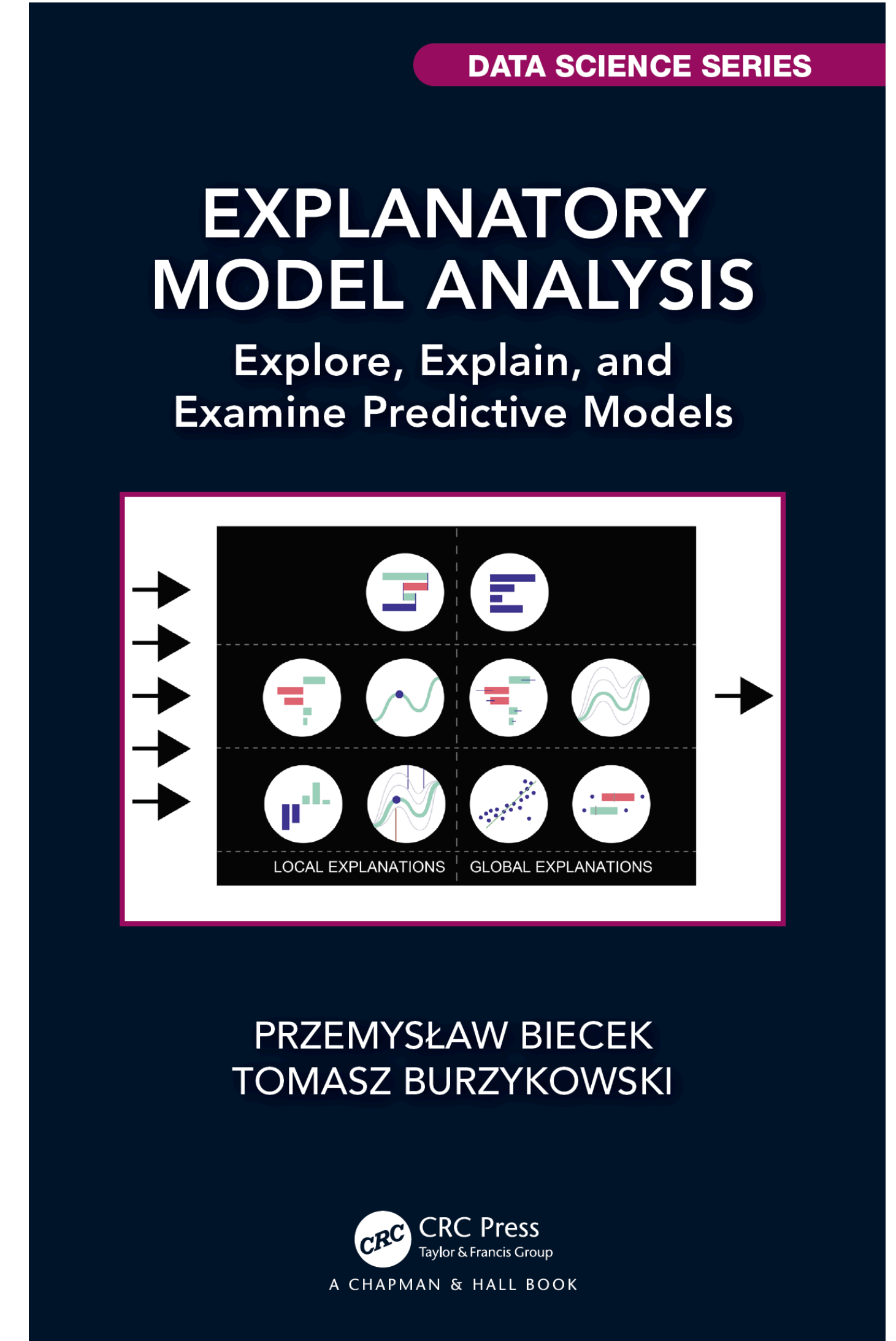
Artı ve eksileri

- Modelin lokal olarak toplamsal olup olmadığını kontrol etmek için kullanılabilirler.
- Model uyumunun yansız olup olmadığını kontrol etmede yararlı olabilirler.

- Karmaşıktırlar.
- Nesnel ölçütler kullanmazlar.

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