

## Giriş

Elektrik, tüketildiği anda üretilmesi gereken bir enerji türüdür. Bu nedenle, üretimi ve tüketimi dengede tutmak önemlidir. Sistem operatörleri, sistemi dengede tutmak için yukarı yönlü düzenlemeler (üretimi artırma/tüketimi azaltma) ve aşağı yönlü düzenlemeler (üretimi azaltma/tüketimi artırma) kullanır. Bir dönem boyunca aşağı yönlü düzenleme ve yukarı yönlü düzenleme arasındaki farka dengesizlik denir. Dengesizlik pozitif veya negatif olabilir. Pozitifse, elektrik sisteminde enerji fazlası olduğu anlamına gelir. Negatifse, enerji açığı olduğu anlamına gelir. Eğer 0 ise, sistem dengededir.

Bu case study'de Jupyter Notebook kullanarak hem markdownlar hem de kod ve çıktılarıyla rapor tarzında bir çalışma tamamladık. Çıktıları, yorumları hem bu raporda hem de Jupyter Notebook dosyasında inceleyebilirsiniz.

Birinci amacımız enerji dengesizliği yönü için sezgisel tahminlerin performansını göstermektir. Bu amaçla, şu adımları takip edeceğiz:

1. EPIAS'tan dengesizlik yönü verilerinin çekilmesi.
2. "Sistem dengesizlik yönü saat t'nin sistem yönü saat t-3 ile aynıdır" şeklindeki sezgisel kuralın performansının değerlendirilmesi.
3. Üç periyot ileri için sistem dengesizlik yönünü tahmin etmek üzere bir makine öğrenimi modeli tasarlanması ve bu modelin performansının değerlendirilmesi

## Veri Toplama

EPIAS'tan dengesizlik yönü verilerini çekmek için Python'un **requests** kütüphanesini kullandık. Bu süreçte, şu URL'yi kullandık:

<https://seffaflik.epias.com.tr/transparency/service/market/smp?endDate=2022-11-01&startDate=2022-01-01>

Bu isteğin yanıtı, her saat için sistem yönü (**smpDirection**) ve tarih (**date**) içeren **body** içinde **smpList** adlı bir JSON formatında verilir. Örneğin, 2022-01-01 00:00 için sistem yönü "YAT" olarak belirtilmiştir. Bu veriyi daha sonra pandas DataFrame formatına dönüştürdük ve çeşitli görselleştirme yöntemleri ile veriyi inceledik.

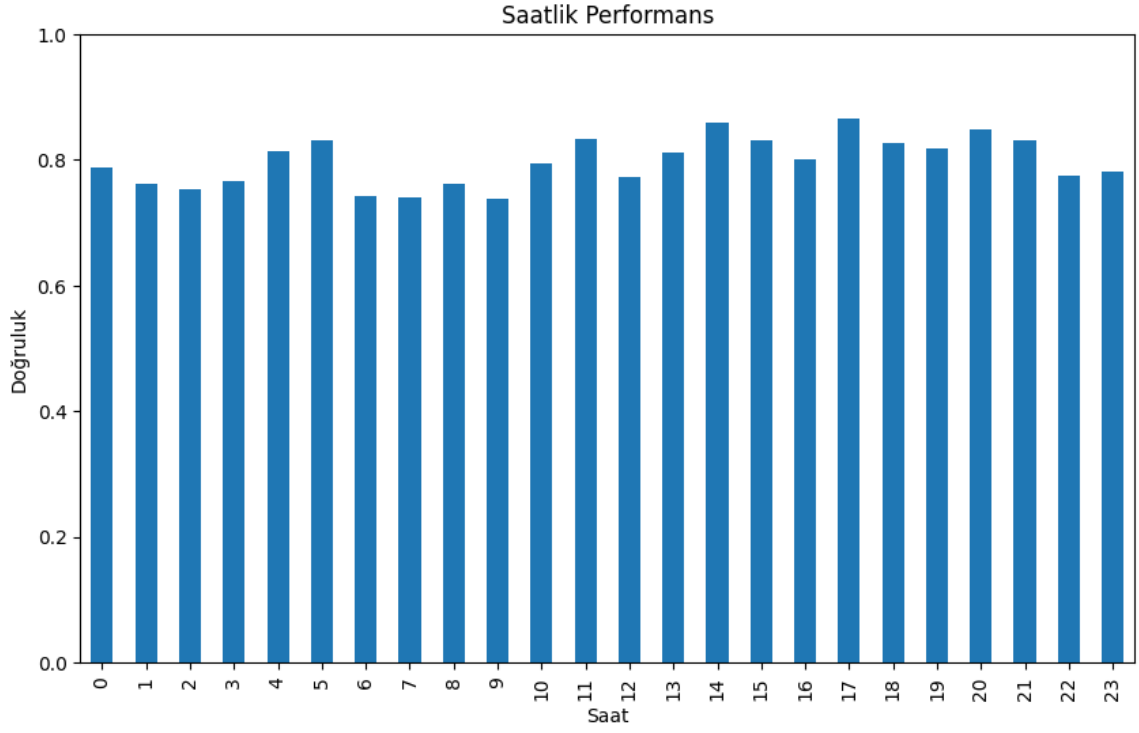
## Heuristik Yöntem

Bu bölümde, t-3 heuristik yöntemini kullanarak sistem dengesizlik yönünü tahmin etmeye çalıştık. Bu yöntem, "Sistem dengesizlik yönü saat t'nin sistem yönü saat t-3 ile aynıdır" şeklindeki sezgisel bir kuralı takip eder. Bu yöntemin performansını değerlendirmek için şu adımları takip ettik:

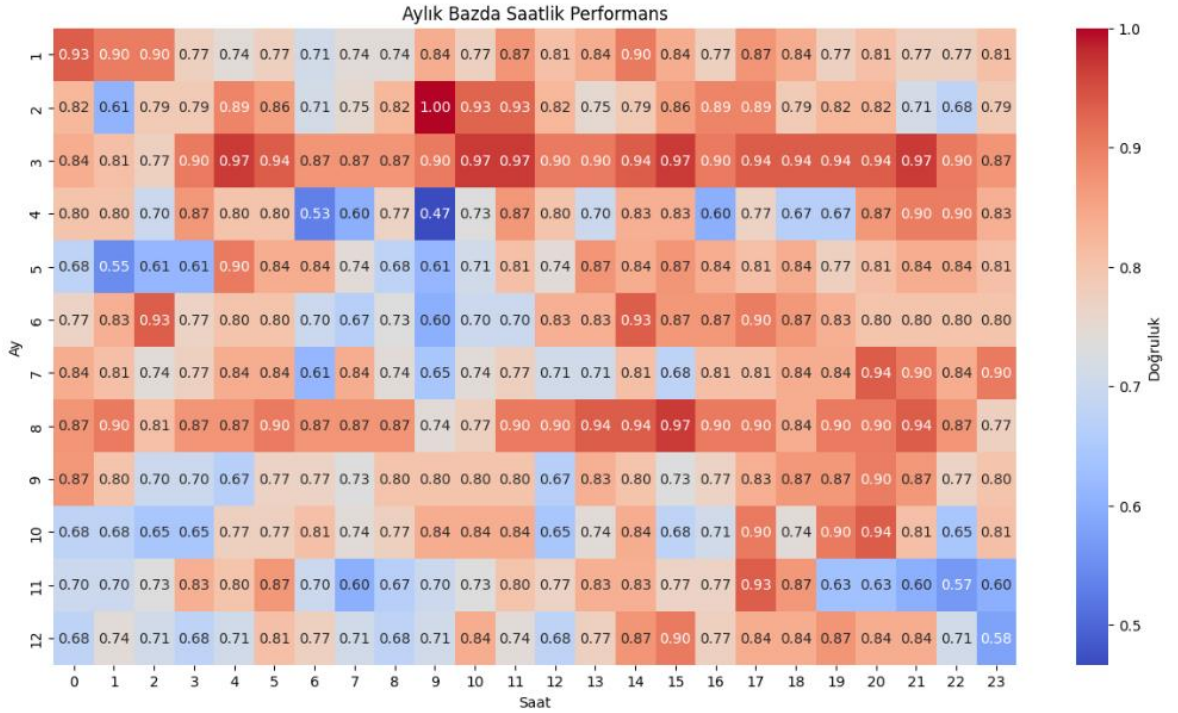
- a) 2022-01-01 ve 2022-12-31 tarihleri arasındaki tüm verileri dikkate alarak heuristik yöntemin doğruluğunu hesapladık.

Sezgisel kuralın performansı: 79.76%

- b) 2022 yılı için her saatlik dilimde heuristik yöntemin doğruluğunu hesapladık.



c) 2022 yılı için her ayın her saatlik diliminde heuristik yöntemin doğruluğunu hesapladık.



d) Bu heuristik yöntemin performansı hakkında yorum yaptık.

### ***Saatlik Performans***

Sezgisel kuralın saatlik performansı grafiđi, kuralın bazı saatlerde daha iyi performans gösterdiğini göstermektedir. Özellikle, saat 17'de en yüksek doğruluk oranına (%86.6) ulaşılırken, saat 9'da en düşük doğruluk oranıyla (%73.7) karşılaşılmaktadır. Bu, enerji talebinin ve arzının belirli saatlerde daha öngörülebilir olduğunu ve kuralın bu saatlerde daha başarılı olduğunu gösteriyor olabilir

### ***Aylık Bazda Saatlik Performansın***

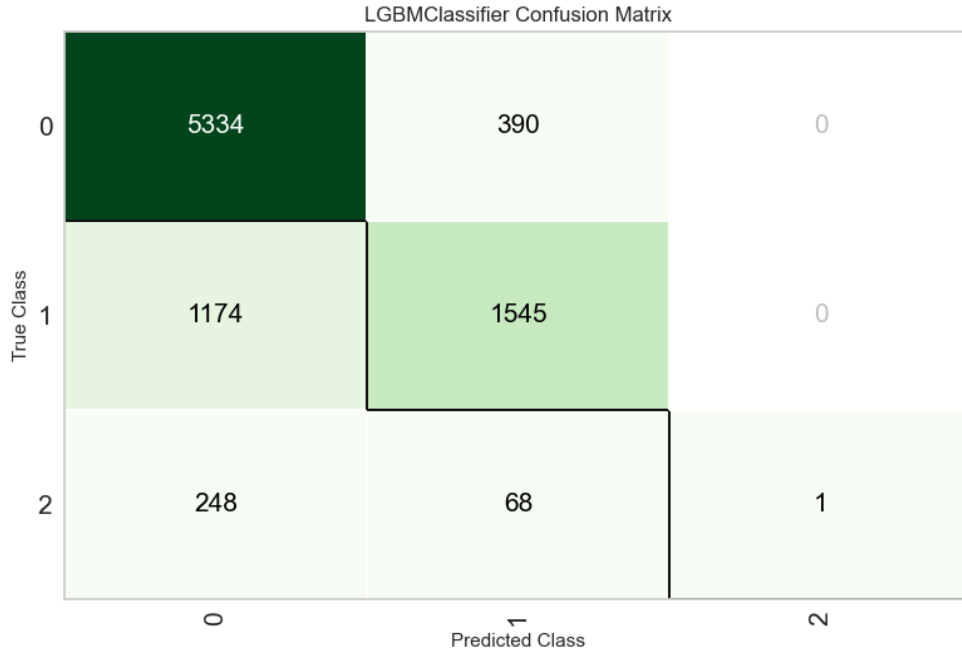
Aylık bazda saatlik performansın ısı haritası, sezgisel kuralın belirli aylarda ve saatlerde nasıl performans gösterdiğini göstermektedir. Özellikle, 2. ayın saat 9'unda %100 doğruluk elde edilirken, 3. ayda ortalama olarak çok iyi tahminler yapılmış ve neredeyse hepsi %90 üzerinde doğruluk sağlanmıştır. Bununla birlikte, 4. ay, sezgisel kuralın en kötü performans gösterdiği dönemdir; özellikle saat 9'da %47 doğruluk elde edilmiştir. Bu bulgular, enerji talebi ve arzının belirli aylar ve saatlerde daha öngörülebilir olduğunu gösteriyor olabilir ve bu durum, sezgisel kuralın başarısını etkileyebilir.

### ***Makine Öğrenimi Modeli***

Bu bölümde, 3 periyot ileri için sistem dengesizlik yönünü tahmin etmek üzere bir makine öğrenimi modeli tasarlamaya çalıştık. Bu amaçla şu adımları takip ettik:

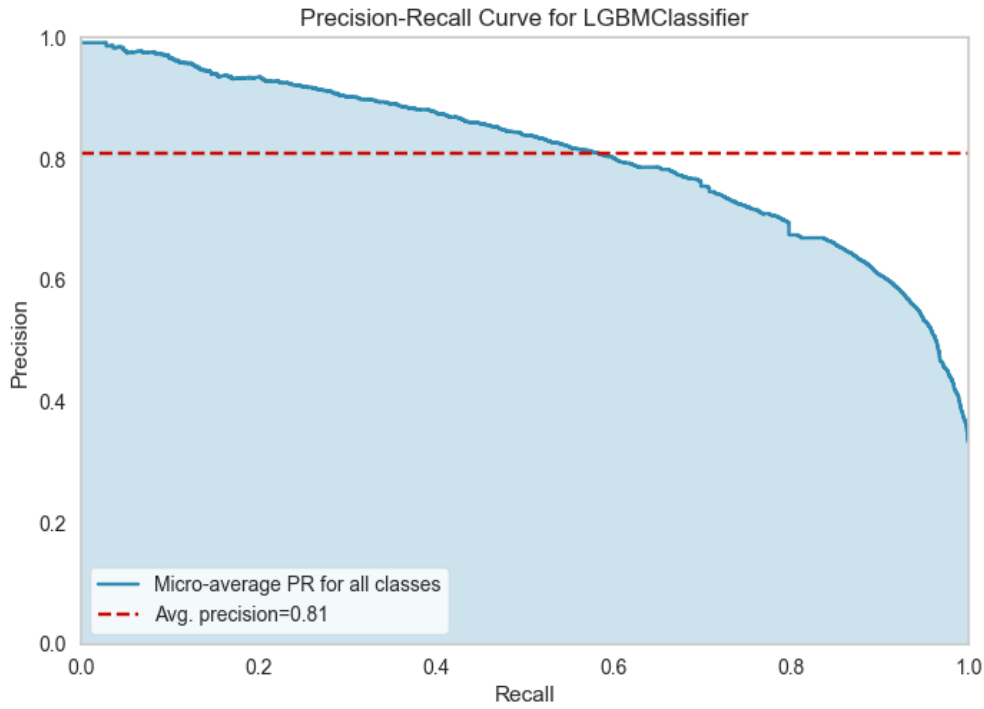
- a) EPIAŞ'tan "DGP Talimat REST SERVİSİ" verilerini çektik.
- b) 3.a'da çektiğimiz verilerden öznitelikler çıkardık ve de farklı yaklaşımlarla birlikte dışsal veriler ekledik. Burada nümerik değişkenler bazında günlük kümülatif ortalama ve 24 saatlik laglı featurelar oluşturduk.
- c) Her periyot için sistem yönünü tahmin etmek üzere bir tahmin modeli tasarladık. PyCaret AutoML kütüphanesi kullanarak bir LightGBM modeli eğittik. Model eğitiminden önce verisetine dışsal veri değişkenleri ekledik (hava durumu, özel günler, covid) daha sonra 291 adetlik verisetinde BorutaPy kütüphanesini ve RandomForestClassifier kullanarak özellik seçimi yaptık. Model performansını çeşitli grafikler kullanarak inceledik ve yorumladık.

### ***Confusion Matrix***



Yukarıda verilen matris, modelin sınıflandırma performansının farklı sınıflar için nasıl olduğunu göstermektedir. Model sınıf 1'i doğru şekilde sınıflandırmada oldukça başarılı olurken, sınıf 2'yi sınıflandırmada daha düşük bir başarı elde eder. Sınıf 3 ise oldukça az sayıda örnek içermesine rağmen, model tarafından sınıflandırılmada en yüksek hata oranına sahiptir.

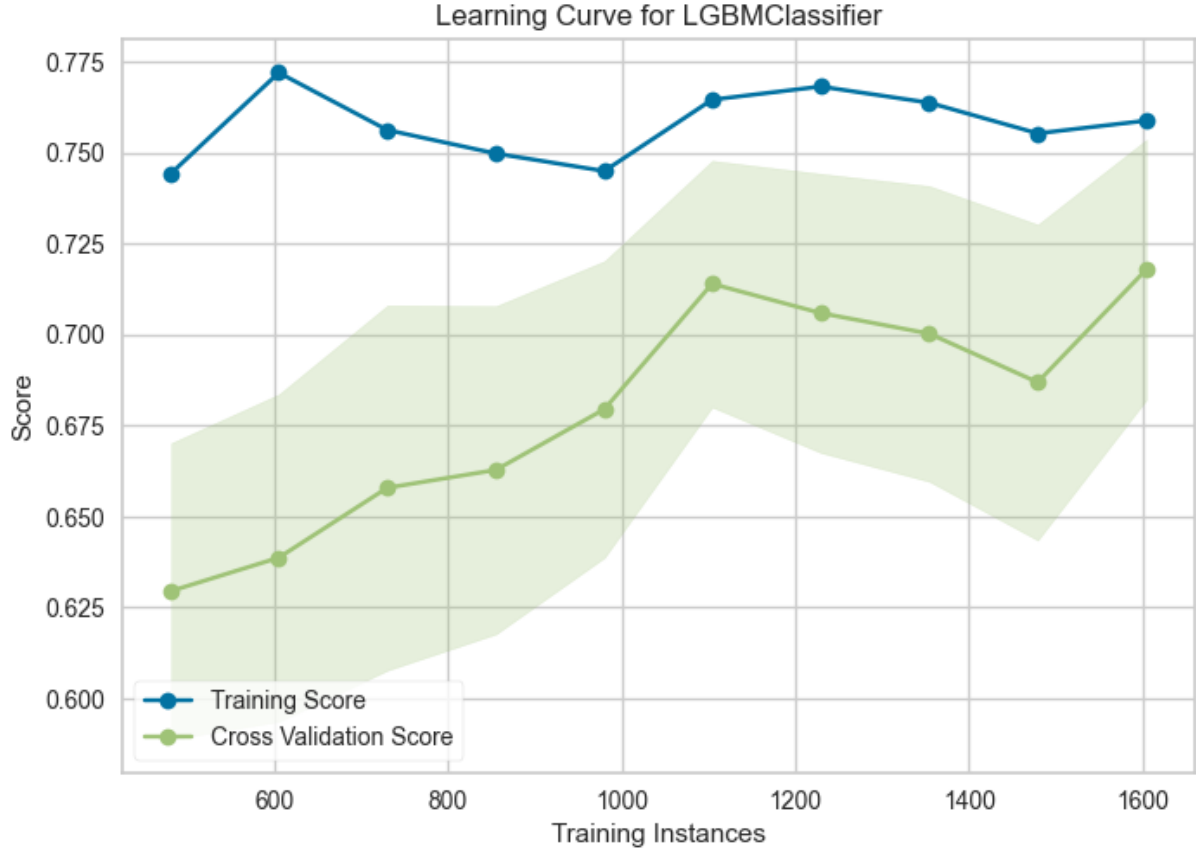
#### PR Eğrisi



Yukarıdaki PR (Precision-Recall) eğrisi, modelin doğru pozitif tahminlerinin sayısını yanıltıcı pozitif tahminlerin sayısına oranını göstererek, modelin sınıflandırma performansının ne kadar iyi olduğunu

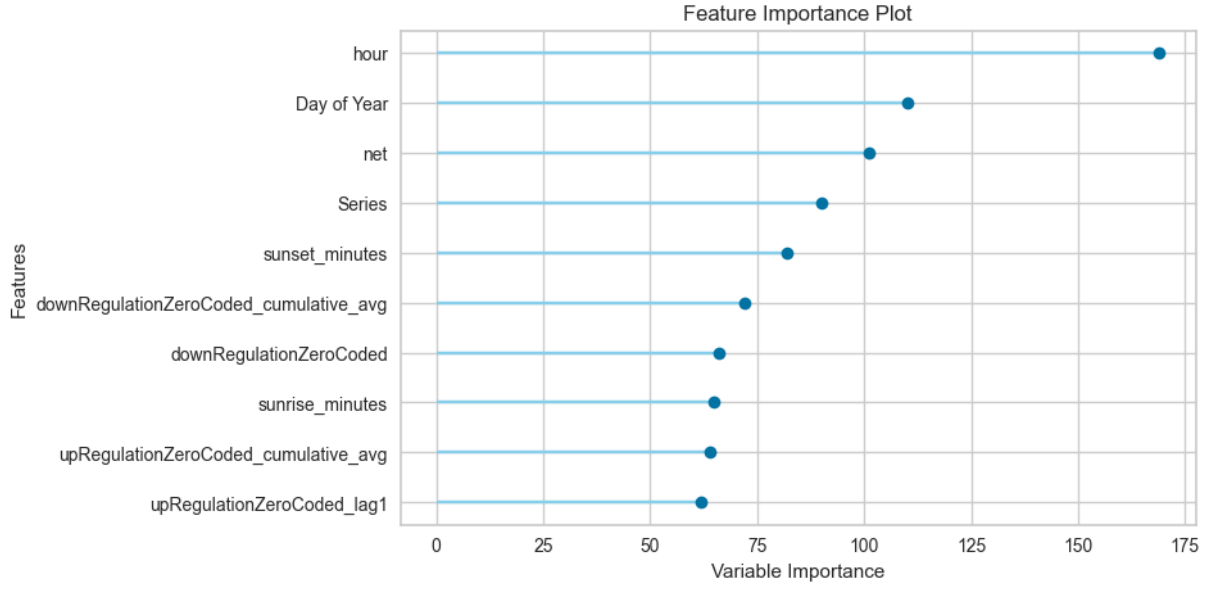
yansıtır. Yüksek bir ortalama hassasiyet değeri, modelin doğru sınıflandırma yapma becerisi olduğunu ve yanıltıcı tahminlerin oranının düşük olduğunu gösterir.

### Learning Curve



Öğrenme eğrisi, başlangıçta düşük örnek sayısı nedeniyle aşırı uyma olasılığı yüksek olduğundan, örnek sayısı arttıkça modelin genelleme yapma becerisinin geliştiğini gösterir. Öğrenme eğrisi, özellikle eğitim skoru ve çapraz doğrulama skoru arasındaki açık farkın daraldığı noktaları vurgulayarak, modelin en iyi performansını hangi noktada gösterdiğini belirleyebilir. Bu durumda, örnek sayısı arttıkça, modelin eğitim ve doğrulama skorları arasındaki fark daralmış ve modelin daha iyi bir genelleme yapma yeteneği olduğu görülmüştür. Fakat buna rağmen geliştirilmeye açık olduğuda gözlemlenmektedir

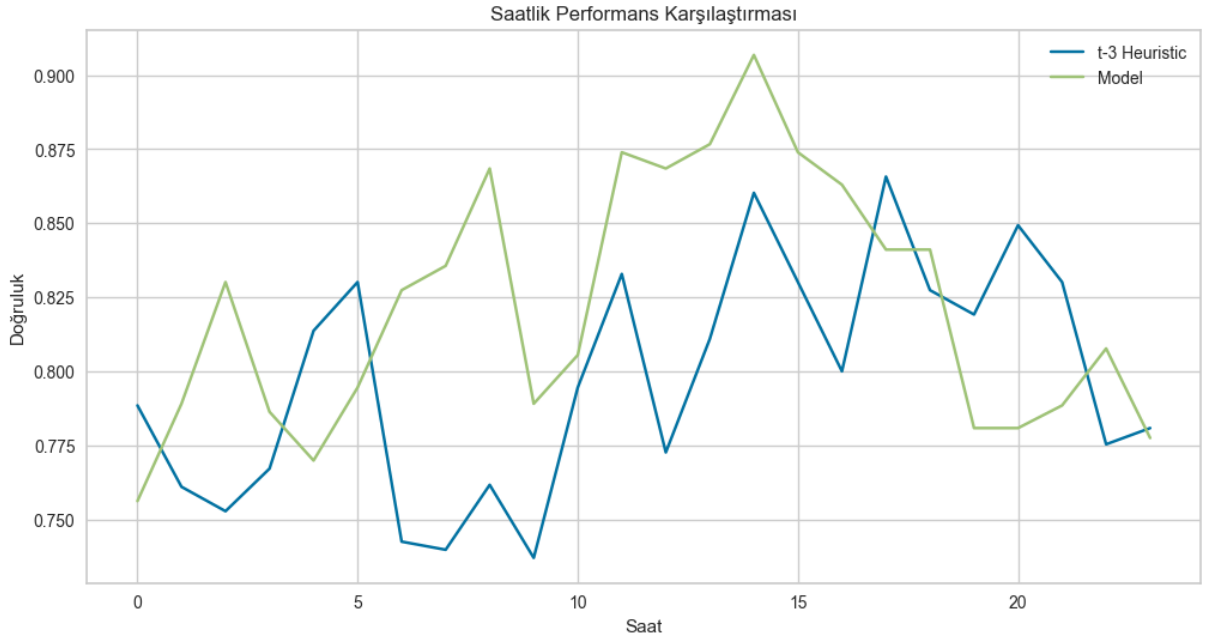
### Özellik Önemi



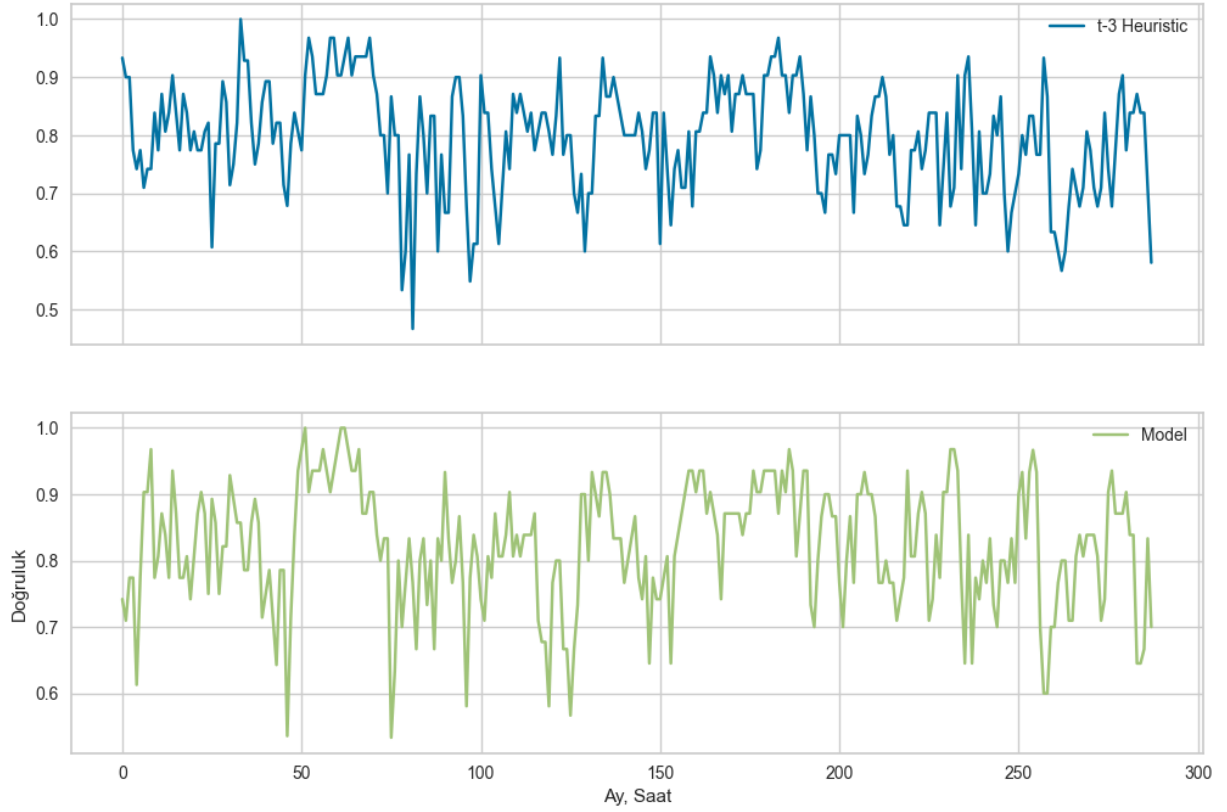
Bu feature importance grafiği, bir makine öğrenimi modelinin hangi özelliklerin (features) sınıflandırmada ne kadar önemli olduğunu gösterir.

Grafiğe göre, bağımlı değişkenin tahmin edilmesinde en önemli özellik saat değişkenidir. Bunu başlangıçtaki sıcaklık grafiğinde de görmüştük. Daha sonra yılın günü değişkeni gelmekte. Üçüncü en önemli değişken ise net değişkenidir. Veri setindeki trendi temsil eden Series değişkeni dördüncü en önemli değişkendir ve devamı yukarıdaki grafikteki gibidir.

d) 2022 verilerini kullanarak test ettik ve 2. bölümdeki heuristik yöntemle sonuçlarımızı karşılaştırdık.



### Aylık Bazda Saatlik Performans Karşılaştırması



Modelin performansını değerlendirdiğimizde, aylık saatlik bazda genel olarak daha iyi sonuçlar elde ettiğini gözlemleyebiliriz. Bu durum, modelin zaman içindeki değişimleri ve dönemsel etkileri daha iyi yakalayabilme yeteneğine işaret ediyor olabilir.

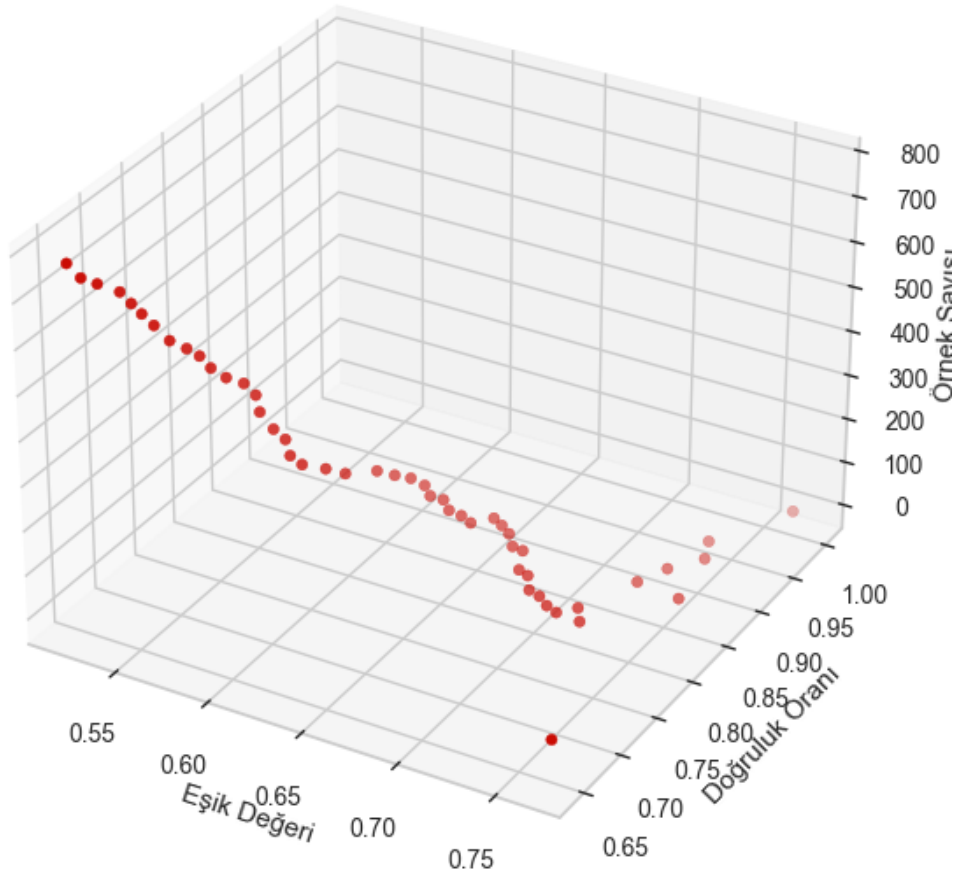
Ayrıca, saatlik bazda modelin 03-05 ve 17-22 saatleri arasında heuristik yöntemle benzer veya biraz daha düşük performans gösterdiğini görüyoruz. Bu saatler dışında ise model, açık ara farkla daha iyi tahminler yapmaktadır. Bu, modelin bu saat aralıklarında belirli özelliklerin etkisinin daha az olduğu veya bu saatlerde enerji piyasalarındaki dengesizliklerin daha öngörülebilir olduğu anlamına gelebilir.

Sonuç olarak, modelimiz saatlik ve aylık saatlik performans açısından başarılı bir şekilde çalışmaktadır. Özellikle belirli saat aralıklarında daha iyi tahminler yaparak enerji piyasalarındaki dengesizlikleri önceden belirlemeye yardımcı olmaktadır.

Özetle, makine öğrenimi modeli, heuristik yöntemle kıyasla daha iyi tahminler yaparak enerji dengesizliği yönünü hem genel, hem saatlik hem de aylık saatlik bazda tahmin etmekte başarılı oldu. Bu, makine öğrenimi modelinin daha karmaşık ilişkileri ve örüntüleri öğrenebilme yeteneğinden kaynaklanmaktadır. Aynı zamanda modelin softmax fonksiyonu kullanarak ürettiği yön tahmini olasılık değerleriyle test verisini kullanarak simülasyon gerçekleştirdik. Bu simülasyon sonucunda modelin ürettiği sınıf olasılıklarının gerçekleşme yüzdeleriyle ilgili analize baktığımızda ilginç sonuçlar elde ettik.

### ***ENERGY\_SURPLUS Sınıfı Tahmin Sonuçları İçin Güven Analizi***

## Doğruluk Oranı, Eşik Değeri ve Örnek Sayısı Arasındaki İlişki



Bu çıktı, modelin enerji fazlası (ENERGY\_SURPLUS) tahminleri için belirli eşik değerlerine göre başarı oranını ve o eşikteki örnek sayısını göstermektedir. Thresh (eşik) sütunu, tahmin olasılığına uygulanan eşik değerini temsil ederken, Yüzde sütunu eşik değerinin üzerindeki tahminlerin doğruluk oranını ve Len sütunu ise bu eşikteki örnek sayısını gösterir.

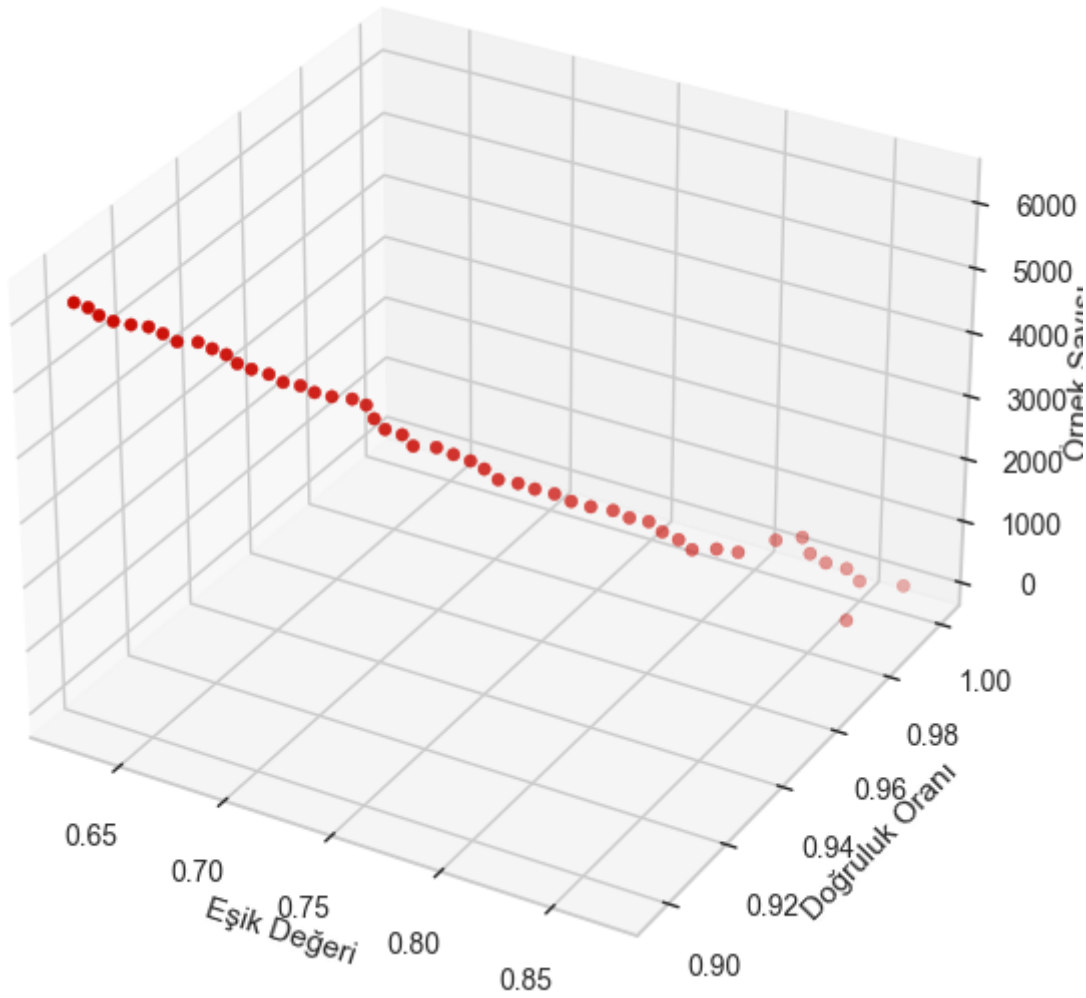
Eşik değeri arttıkça, modelin daha yüksek "prediction\_score" değerlerine sahip örnekleri seçtiği ve dolayısıyla doğruluk oranının da arttığı görülmektedir. Örneğin, 0.765 eşik değerinde sadece 1 örnek bulunmakta ve bu örnek için doğruluk oranı %100'dür. Ancak, eşik değeri düştükçe, örnek sayısı artmakta ve doğruluk oranı genellikle azalmaktadır.

Eşik değeri 0.520'de ise doğruluk oranı %64.4 ve örnek sayısı 772'dir. Bu durum, eşik değerlerini düşürdükçe modelin doğruluk oranının azaldığını ve daha fazla örnekleme tabi olduğunu gösterir. Enerji dengesizliği yön tahminlerinde, doğru bir denge bulmak önemlidir, çünkü yüksek doğruluk oranları elde etmek istiyoruz, ancak aynı zamanda yeterli örnek sayısına sahip olmak da önemlidir.



Sonuç olarak, bu çıktılar modelin "ENERGY\_SURPLUS" tahminleri için farklı eşik değerlerinde nasıl performans gösterdiğini anlamamıza yardımcı olmaktadır. Bu bilgi, enerji sektöründe karar vericilere, modelin performansını ve güvenilirliğini değerlendirmelerine ve en uygun eşik değerini seçmelerine yardımcı olabilir.

#### **ENERGY\_DEFICIT Sınıfı Tahmin Sonuçları İçin Güven Analizi**



Bu çıktı, modelin enerji fazlası (ENERGY\_SURPLUS) tahminleri için belirli eşik değerlerine göre başarı oranını ve o eşikteki örnek sayısını göstermektedir. Thresh (eşik) sütunu, tahmin olasılığına uygulanan eşik değerini temsil ederken, Yüzde sütunu eşik değerinin üzerindeki tahminlerin doğruluk oranını ve Len sütunu ise bu eşikteki örnek sayısını gösterir.

Tabloya göre, eşik değeri yükseldikçe modelin doğruluk oranı genel olarak artmaktadır. Örneğin, 0.870 eşik değeri için modelin başarısı %100'dür, ancak bu durumda sadece 76 örnek dikkate alınmıştır. Eşik değeri düştükçe, örnek sayısı artar ve modelin başarısı azalır. Bu, düşük eşik

değerlerinin daha geniş bir örnekleme kapsadığını ve modelin daha yüksek olasılıklar için daha güvenilir tahminler ürettiğini gösterir.

Öte yandan, eşik değerlerinin çok yüksek seçilmesi, modelin sadece çok yüksek güvene sahip olduğu örnekler için doğru tahminler yapabileceği anlamına gelir. Bu durum, tahminlerin gerçekleşme olasılığının daha düşük olduğu durumlarda modelin başarısını sınırlayabilir. Bu nedenle, doğru eşik değerini seçmek, doğru tahminler ile örnek sayısı arasında dikkatli bir denge kurmayı gerektirir.

## **Sonuç**

Bu case study'de, enerji dengesizliği yönünün tahmin edilmesi için hem heuristik bir yöntem hem de makine öğrenimi modeli kullandık. Heuristik yöntem, makul bir doğruluk seviyesi sağlasa da, makine öğrenimi modeli daha iyi tahminler yaparak enerji dengesizliği yönünü tahmin etmekte başarılı oldu.

Makine öğrenimi modelinin performansı, verilerin özelliklerini ve kullanılan algoritmanın doğasını dikkate alarak daha da geliştirilebilir. Ayrıca, farklı saatlerde ve aylarda performansı inceleyerek, modelin belirli zaman dilimlerinde nasıl iyileştirilebileceğine dair önerilerde bulunabiliriz.

Bu çalışma, enerji dengesizliği yönünün tahmin edilmesinde heuristik yöntemlerin ve makine öğrenimi modellerinin kullanılabilirliğini ve etkinliğini göstermektedir. Bu tür tahminler, enerji sektöründe karar vericilere ve enerji tedarikçilerine, enerji dengesizliklerini yönetmek ve enerji sistemlerini daha verimli hale getirmek için önemli bilgiler sunar.