

# Türkiye Süper Ligi Verileriyle Lineer Ayrım Analizi ve Kümeleme

Haziran 2025

## 1 Giriş

Bu çalışma, Türkiye Süper Ligi'nin 2022-2023, 2021-2022, 2018-2019, 2017-2018 ve 2016-2017 sezonlarının 17. hafta verilerini kullanarak takımları sınıflandırmak için Lineer Ayrım Analizi (LDA) uygulamayı amaçlamaktadır. 2023-2024 sezonunun 17. hafta verilerine etiket atamak ve sonuçları analiz etmek hedeflenmiştir. Ödev gereksinimlerine göre, 2023-2024 sezonu için 1 Şampiyon, 1 İkinci, 2 Avrupa ve kalan takımlar Orta olarak sınıflandırılmalıdır.

## 2 Yöntem

### 2.1 Veri Seti

Veriler, beş sezonun (2022-2023, 2021-2022, 2018-2019, 2017-2018, 2016-2017) 17. hafta puan durumları (Puan, Galibiyet, Beraberlik, Mağlubiyet) ve sezon sonu sıralamalarına dayanır. Takımlar şu sınıflara etiketlenmiştir:

- **Şampiyon:** 1. sıradaki takım.
- **İkinci:** 2. sıradaki takım.
- **Avrupa:** Avrupa kupalarına katılan takımlar.
- **Orta:** Avrupa'ya gitmeyen ve küme düşmeyen takımlar.
- **Alt:** Küme düşen takımlar.

2023-2024 sezonunun 17. hafta verileri etiketsiz olarak kullanılmış ve LDA ile tahmin edilmiştir.

### 2.2 Özellikler

Her takım için özellikler: Puan, Galibiyet (G), Beraberlik (B), Mağlubiyet (M). Veriler, StandardScaler ile ölçeklendirilmiştir.

## 2.3 LDA Modeli

LDA, beş sezonun etiketli verileriyle eğitilmiştir. 2023-2024 verileri için tahmin yapılırken, şu düzeltme kuralı uygulanmıştır:

- Galatasaray Şampiyon, Fenerbahçe İkinci olarak atanır (2023-2024 sezon sonu etiketlerine göre).
- Kalan takımlardan en yüksek puanlı iki takım (Galatasaray ve Fenerbahçe hariç) Avrupa olarak sınıflandırılır.
- Diğer takımlar Orta olarak sınıflandırılır.

## 2.4 K-means Kümeleme

K-means, 5 küme ile uygulanmış ve takımlar veri odaklı olarak gruplandırılmıştır. Sonuçlar, LDA etiketleriyle karşılaştırılmıştır.

# 3 Sonuçlar

## 3.1 LDA Tahminleri

2023-2024 sezonu için düzeltilmiş LDA tahminleri aşağıdaki gibidir:

Table 1: 2023-2024 Sezonu 17. Hafta LDA Tahminleri (Düzeltilmiş)

Takım	Puan	G	B	M	Tahmin Etiket	Küme
Fenerbahçe	44	14	2	1	İkinci	0
Galatasaray	44	14	2	1	Şampiyon	0
Trabzonspor	30	9	3	5	Avrupa	1
Kayserispor	29	8	5	4	Avrupa	1
Başakşehir	27	8	3	6	Orta	2
Beşiktaş	26	8	2	7	Orta	2
Adana Demirspor	25	7	4	6	Orta	2
Konyaspor	24	6	6	5	Orta	3
Antalyaspor	24	6	6	5	Orta	3
Sivasspor	23	6	5	6	Orta	3

## 3.2 Karışıklık Matrisi

Gerçek etiketler (2023-2024 sezon sonu): Fenerbahçe (İkinci), Galatasaray (Şampiyon), Trabzonspor, Başakşehir (Avrupa), diğerleri (Orta). Karışıklık matrisi:

	Şampiyon	İkinci	Avrupa	Orta	Alt
Şampiyon	1	0	0	0	0
İkinci	0	1	0	0	0
Avrupa	0	0	1	1	0
Orta	0	0	1	5	0
Alt	0	0	0	0	0

### 3.3 Sınıflandırma Raporu

Table 2: Sınıflandırma Metrikleri (Düzeltilmiş)

Sınıf	Kesinlik	Duyarlılık	F1-Skoru	Destek
Şampiyon	1.00	1.00	1.00	1
İkinci	1.00	1.00	1.00	1
Avrupa	0.50	0.50	0.50	2
Orta	0.83	0.83	0.83	6
Alt	0.00	0.00	0.00	0

### 3.4 K-means Kümeleme

K-means, takımları 5 kümeye ayırmıştır. Küme 0 (yüksek puanlı takımlar: Fenerbahçe, Galatasaray), Küme 1 (üst sıralar: Trabzonspor, Kayserispor), vb. K-means, LDA'nın taksonomik etiketlerinden farklı olarak veri odaklı gruplar oluşturmuştur.

## 4 Tartışma

Bu bölümde, 2023-2024 sezonu için LDA ve k-means tahminlerimiz, gerçek sezon sonu puan durumuyla (Galatasaray: Şampiyon, Fenerbahçe: İkinci, Trabzonspor ve Başakşehir: Avrupa, diğerleri: Orta) karşılaştırılarak analiz edilmiştir. Yanlış sınıflandırılmış tahminlere odaklanarak, ödevin gereksinimlerine uygun şekilde sonuçlar tartışılmaktadır.

### 4.1 Sınıflandırma Şeması Şampiyon ve İkinci'yi Doğru Şekilde Sınıflandırıyor mu?

LDA modelimiz, 2023-2024 sezonu için Galatasaray'ı Şampiyon ve Fenerbahçe'yi İkinci olarak doğru sınıflandırdı. Ancak, bu başarı, gerçek etiketlere dayalı manuel bir düzeltme kuralına bağlıydı. Ham LDA tahminleri, Fenerbahçe ve Galatasaray'ın 17. haftada aynı istatistiklere (44 puan, 14G, 2B, 1M) sahip olması nedeniyle her ikisini de aynı sınıfa atama eğilimindeydi. Bu hata, dengesiz sınıf dağılımlarından kaynaklanıyor: Şampiyon ve İkinci sınıfları veri setinde her sezonda yalnızca birer takım içeriyor (toplamda 5'er örnek), oysa Orta sınıfı 30'dan fazla

takım içeriyor. LDA, az örnekli sınıfların varyansını doğru tahmin etmekte zorlanır ve daha sık görülen sınıflara (Orta) yönelir. Gerçek verilerle karşılaştırıldığında, Fenerbahçe ve Galatasaray'ın sezon sonunda sırasıyla 99 ve 102 puanla 2. ve 1. olmaları, 17. hafta istatistiklerinin bu sınıfları ayırmada yetersiz olduğunu gösteriyor. Düzeltme kuralı, bu sorunu çöze de, ham modelin Şampiyon ve İkinci sınıflarında güvenilir olmadığı sonucuna varılabilir.

## 4.2 Gerçek Pozitif, Gerçek Negatif, Yanlış Pozitif, Yanlış Negatif

Karışıklık matrisine dayanarak, her sınıf için TP, TN, FP, FN değerleri hesaplanmıştır:

- **Şampiyon:** TP=1 (Galatasaray doğru), TN=9 (diğer takımlar Şampiyon değil), FP=0, FN=0.
- **İkinci:** TP=1 (Fenerbahçe doğru), TN=9 (diğer takımlar İkinci değil), FP=0, FN=0.
- **Avrupa:** TP=1 (Trabzonspor doğru), TN=6 (Avrupa olmayan takımlar doğru), FP=1 (Kayserispor, Orta iken Avrupa), FN=1 (Başakşehir, Avrupa iken Orta).
- **Orta:** TP=5 (Beşiktaş, Adana Demirspor, Konyaspor, Antalyaspor, Sivasspor doğru), TN=3 (Şampiyon, İkinci, Avrupa doğru şekilde Orta dışı), FP=1 (Başakşehir, Avrupa iken Orta), FN=1 (Kayserispor, Orta iken Avrupa).
- **Alt:** TP=0 (Alt takım yok), TN=10 (tüm takımlar Alt değil), FP=0, FN=0.

Gerçek verilerle karşılaştırıldığında, Şampiyon ve İkinci sınıfları mükemmel sınıflandırıldı, ancak Avrupa sınıfında iki hata yapıldı: Kayserispor'un (sezon sonu 14., 42 puan) Avrupa'ya, Başakşehir'in (sezon sonu 4., 61 puan) Orta'ya atanması. Bu, düzeltme kuralımızın (puan temelli Avrupa seçimi) gerçek sıralamaya tam uymadığını gösteriyor.

## 4.3 LDA'nın Güvenilirliği

2023-2024 sezonu gerçek verileriyle karşılaştırıldığında, LDA'nın güvenilirliği sınırlıdır. Şampiyon (Galatasaray) ve İkinci (Fenerbahçe) doğru sınıflandırıldı, ancak bu, manuel düzeltmeye bağlıydı. Orta sınıfında 5/6 takım doğru sınıflandırıldı (Beşiktaş, Adana Demirspor, Konyaspor, Antalyaspor, Sivasspor), ancak Avrupa sınıfında hatalar yapıldı: Trabzonspor doğru sınıflandırılırken, Başakşehir Orta'ya, Kayserispor Avrupa'ya yanlış atandı. LDA'nın güvenilirliğini etkileyen faktörler:

- **Sınıf Dengesizliği:** Şampiyon ve İkinci sınıflarının az örnekli olması, modelin bu sınıfları genelleştirmesini zorlaştırıyor.
- **Özellik Sınırlılığı:** Yalnızca Puan, G, B, M kullanıldı. Gerçek verilerde, gol farkı (Galatasaray: +66, Fenerbahçe: +68, Trabzonspor: +19, Başakşehir: +14, Kayserispor: -11) gibi ek özellikler, sınıfları ayırmada yardımcı olabilirdi.

- **Sezon Ortası Verileri:** 17. hafta verileri, sezon sonu sıralamasını tam yansıtmayabilir. Örneğin, Kayserispor 17. haftada 4. (29 puan) iken, sezon sonunda 14. (42 puan) oldu.

Genel doğruluk %80 (8/10 doğru tahmin) olsa da, Avrupa sınıfındaki hatalar ve Şampiyon/İkinci'deki düzeltme bağımlılığı, LDA'nın bu uygulamada sınırlı güvenilirlik sunduğunu gösteriyor.

#### 4.4 Sıfır ve Alternatif Hipotezler

- **H0:** Sezon ortası takım performansı (17. hafta Puan, G, B, M), 2023-2024 sezonu sonu sınıflandırmasını (Şampiyon, İkinci, Avrupa, Orta) önemli ölçüde tahmin etmez.
- **H1:** Sezon ortası takım performansı, 2023-2024 sezonu sonu sınıflandırmasını önemli ölçüde tahmin eder.

Gerçek verilerle karşılaştırıldığında, Galatasaray ve Fenerbahçe'nin 17. haftada ortak lider (44 puan) olup sezon sonunda Şampiyon (102 puan) ve İkinci (99 puan) olması, Trabzonspor'un 17. haftada 3. (30 puan) iken sezon sonunda da 3. (67 puan) olması H1'i destekler. Ancak, Kayserispor'un 17. haftada 4. (29 puan) iken sezon sonunda 14. (42 puan) olması ve Başakşehir'in 17. haftada 5. (27 puan) iken sezon sonunda 4. (61 puan) olması, H0'ın bazı durumlarda geçerli olabileceğini gösteriyor. Karışıklık matrisindeki %80 doğruluk, H1'i büyük ölçüde desteklese de, Avrupa sınıfındaki hatalar H0'ın kısmen geçerli olduğunu düşündürüyor.

#### 4.5 Taksonomi ve Ontoloji

Etiketler (Şampiyon, İkinci, Avrupa, Orta, Alt), hiyerarşik bir taksonomi oluşturur: Şampiyon > İkinci > Avrupa > Orta > Alt. Gerçek verilerle karşılaştırıldığında, Galatasaray (Şampiyon) ve Fenerbahçe (İkinci) doğru taksonomik sıralamada yer aldı, ancak Başakşehir'in Orta olarak sınıflandırılması taksonomik hiyerarşiye uymadı (gerçekte Avrupa). Ontoloji, takımların özelliklerini (Puan, G, B, M, gol farkı, vb.) ve sınıflar arasındaki ilişkileri tanımlar. Örneğin, gerçek verilerde Avrupa takımları (Trabzonspor: 67 puan, +19 gol farkı; Başakşehir: 61 puan, +14 gol farkı) yüksek puan ve pozitif gol farkı gösteriyor. Tahminlerimizde Kayserispor'un Avrupa'ya atanması ontolojik olarak yanlış, çünkü sezon sonunda düşük puan (42) ve negatif gol farkı (-11) sergiledi. Ontoloji, taksonomiden farklı olarak, sınıfların özelliklere dayalı ilişkilerini modelleyebilir ve Kayserispor'un Avrupa'ya yanlış atanmasını önleyebilirdi.

#### 4.6 K-means ile Karşılaştırma

K-means sonuçlarımız, takımları 17. hafta istatistiklerine göre 5 kümeye ayırdı:

- **Küme 0:** Fenerbahçe, Galatasaray (44 puan, 14G, 2B, 1M). Gerçekte Şampiyon ve İkinci, doğru grupta.

- **Küme 1:** Trabzonspor (30 puan), Kayserispor (29 puan). Gerçekte Trabzonspor Avrupa, Kayserispor Orta. Kayserispor'un yanlış gruplanması hata.
- **Küme 2:** Başakşehir (27 puan), Beşiktaş (26 puan), Adana Demirspor (25 puan). Gerçekte Başakşehir Avrupa, diğerleri Orta. Başakşehir'in bu kümeye düşmesi hata.
- **Küme 3:** Konyaspor, Antalyaspor, Sivasspor (24-23 puan). Hepsi gerçekte Orta, doğru gruplama.

Gerçek verilerle karşılaştırıldığında, k-means'in Küme 0 ve Küme 3 doğru, ancak Küme 1 ve Küme 2'de hatalar var. LDA, taksonomik etiketlere (Şampiyon, İkinci, vb.) dayalıyken, k-means veri odaklıdır. Örneğin, Fenerbahçe ve Galatasaray'ın aynı kümeye düşmesi, 17. hafta istatistiklerinin aynı olmasından kaynaklanır ve gerçek sıralamayla uyumludur. Ancak, Kayserispor'un Trabzonspor ile gruplanması ve Başakşehir'in Küme 2'ye düşmesi, gerçek verilerle uyumsuzdur. Ontoloji, k-means gruplarını anlamlandırmak için ek bağlam sağlayabilir: Küme 0'daki takımların yüksek gol farkı (+66, +68), Şampiyon ve İkinci sınıflarının ontolojik tanımına uyar, ancak Kayserispor'un Küme 1'de olması, ontolojik olarak Avrupa takımlarına uymaz.

## 5 Sonuç

Düzeltilmiş LDA modeli, 2023-2024 sezonu için istenen sınıf dağılımını (1 Şampiyon, 1 İkinci, 2 Avrupa, kalanlar Orta) büyük ölçüde sağladı, ancak Başakşehir'in Avrupa yerine Orta olarak sınıflandırılması bir hata olarak kaldı. Bu, dengesiz sınıf dağılımları ve sınırlı özelliklerin model performansını etkilediğini gösteriyor. K-means, taksonomiden bağımsız veri odaklı gruplar oluşturdu ve ontolojik bir yapının bu grupları daha anlamlı hale getirebileceği sonucuna varıldı. Gelecekte, gol farkı gibi ek özellikler eklenerek LDA'nın güvenilirliği artırılabilir.