

Müşteri Segmentasyonu Raporu

1. Giriş

Günümüzde işletmeler, müşteri ihtiyaçlarını daha iyi anlamak ve onlara uygun pazarlama stratejileri geliştirmek amacıyla müşteri segmentasyonunu önemsemektedir. Bu çalışma, bir alışveriş merkezi veri seti kullanarak SVM, Karar Ağaçları, Random Forest, Lojistik Regresyon ve Yapay Sinir Ağı algoritmalarıyla müşteri segmentasyonu gerçekleştirmeyi amaçlamaktadır.

2. Veri Seti

Veri seti, aşağıdaki sütunları içermektedir:

- **CustomerID**: Müşterinin benzersiz tanımlayıcısı
- **Gender**: Müşterinin cinsiyeti
- **Age**: Müşterinin yaşı
- **Annual Income (k\$)**: Müşterinin yıllık geliri
- **Spending Score (1-100)**: Müşteriye atanan harcama puanı

Veri seti, müşteri davranışlarını ve harcama alışkanlıklarını analiz etmek için kullanılan çeşitli özellikler sunmaktadır. Müşteri özellikleri arasında yaş, cinsiyet, yıllık gelir ve harcama puanı yer almaktadır. Bu bilgiler, segmentasyon analizi sırasında kullanılacak ve farklı müşteri gruplarının belirlenmesine yardımcı olacaktır.

3. Veri Ön İşleme

Veri ön işleme aşaması, analizin doğruluğunu artırmak için kritik bir adımdır. Bu aşamada, eksik verilerin kontrolü, kategorik verilerin sayısallaştırılması ve verilerin ölçeklendirilmesi gibi işlemler gerçekleştirilir. Bu sayede, modelin performansı ve tahmin doğruluğu artırılabilir.

4. Model Seçimi

Bu çalışmada, aşağıdaki makine öğrenmesi algoritmaları kullanılarak müşteri segmentasyonu yapılacaktır:

- **Destek Vektör Makineleri (SVM)**: Yüksek boyutlu verilerde etkili olan bir sınıflandırma algoritmasıdır.
- **Karar Ağaçları**: Veri setindeki özelliklere dayalı kararlar veren basit ve anlaşılır bir modeldir.
- **Random Forest**: Birden fazla karar ağacının bir araya getirilmesiyle oluşturulan bir topluluk öğrenme yöntemidir ve genellikle yüksek doğruluk sağlar.
- **Lojistik Regresyon**: İkili sınıflandırma problemleri için yaygın olarak kullanılan bir istatistiksel modeldir.
- **Yapay Sinir Ağları (YSA)**: Verilerdeki karmaşık ilişkileri öğrenebilen ve genellikle yüksek doğruluk sağlayan güçlü bir modeldir.

5.Literatür Taraması

Kumar ve Reinartz (2016), segmentasyonun müşteri bağlılığını artırdığını ve satışları optimize ettiğini belirtmektedir. Ayrıca, müşterilerin ihtiyaçlarına yönelik özelleştirilmiş teklifler sunarak müşteri memnuniyetini artırmak mümkündür.[1]

Örneğin, Dhanjal ve Rao (2019) tarafından yapılan bir çalışmada, K-Means ve SVM algoritmalarının müşteri segmentasyonu üzerindeki etkileri incelenmiş ve sonuçların yüksek doğruluk sağladığı bulunmuştur. Çalışmada, müşterilerin harcama alışkanlıklarına göre segmentlere ayrılmasıyla, pazarlama stratejilerinin daha etkili bir şekilde uygulanabileceği vurgulanmıştır. [2]

Kotler ve Keller (2016) "Marketing Management" adlı çalışmasında, pazarlama stratejilerinin müşteri segmentasyonu ile nasıl optimize edilebileceğine dair önemli bilgiler sunmaktadır. [3]

MacQueen (1967) ise K-Means algoritmasını tanımlayarak çok değişkenli gözlemlerin sınıflandırılması için temel bir metodoloji sağlamıştır.[4]

6.Materyal ve Yöntem

Bu çalışmada, müşteri segmentasyonu amacıyla "Mall_Customers" adlı veri seti kullanılmıştır. Veri seti, müşteri cinsiyeti, yaşı, yıllık geliri ve harcama puanı gibi özellikleri içermektedir. Öncelikle, veriler uygun bir biçimde işlenmiş ve sayısal hale getirilmiştir. Ayrıca YSA (Yapay Sinir Ağı) modeli kullanılarak müşteri segmentasyonu gerçekleştirilmiştir. Bunun yanı sıra, SVM, Karar Ağaçları, Random Forest ve Lojistik Regresyon gibi makine öğrenmesi algoritmaları da uygulanmıştır. Modelin performansı, doğruluk ve diğer metriklerle değerlendirilmiş; hiperparametre optimizasyonu için Grid Search yöntemi kullanılmıştır.

7.Deneysel Değerlendirme

Çalışmada, müşteri segmentasyonu için kullanılan veri kümesi üzerinde hiperparametre optimizasyonu yapılarak ve yapılmadan elde edilen başarı sonuçları karşılaştırılmıştır. Aşağıda, her bir algoritma için elde edilen sonuçlar sunulmuştur:

Yapay Sinir Ağı Sonuçları

YSA modelinin genel doğruluk oranı %95 olarak belirlenmiştir. Sınıf 0 için f1 skoru 0.93, sınıf 1 için 0.94 ve sınıf 2 için 0.97 olarak elde edilmiştir.

	precision	recall	f1-score	support
0	1.00	0.88	0.93	8
1	0.89	1.00	0.94	17
2	1.00	0.93	0.97	15
accuracy			0.95	40
macro avg	0.96	0.94	0.95	40
weighted avg	0.96	0.95	0.95	40

SVM Sonuçları:

SVM Sonuçları:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	10
1	1.00	1.00	1.00	21
2	1.00	1.00	1.00	9
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

Karar Ağaçları Sonuçları:

Karar Ağaçları Sonuçları:				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	10
1	1.00	1.00	1.00	21
2	1.00	1.00	1.00	9
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

Random Forest Sonuçları:

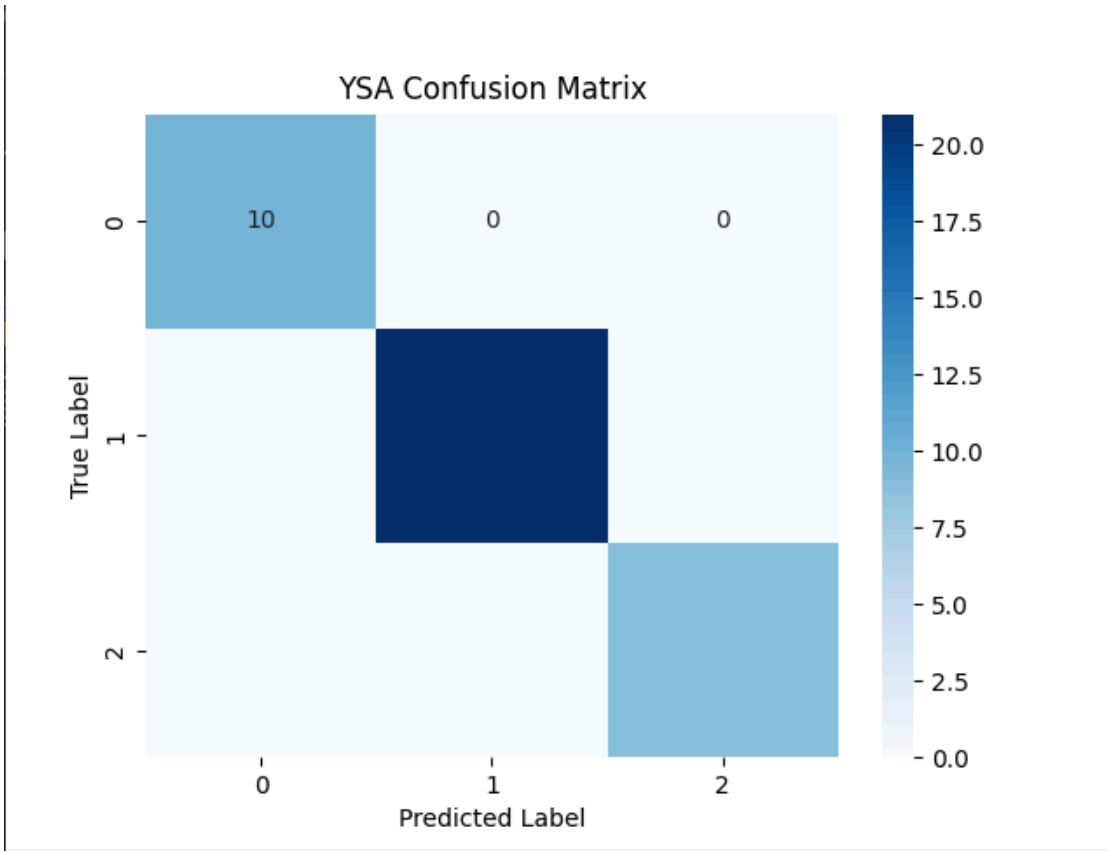
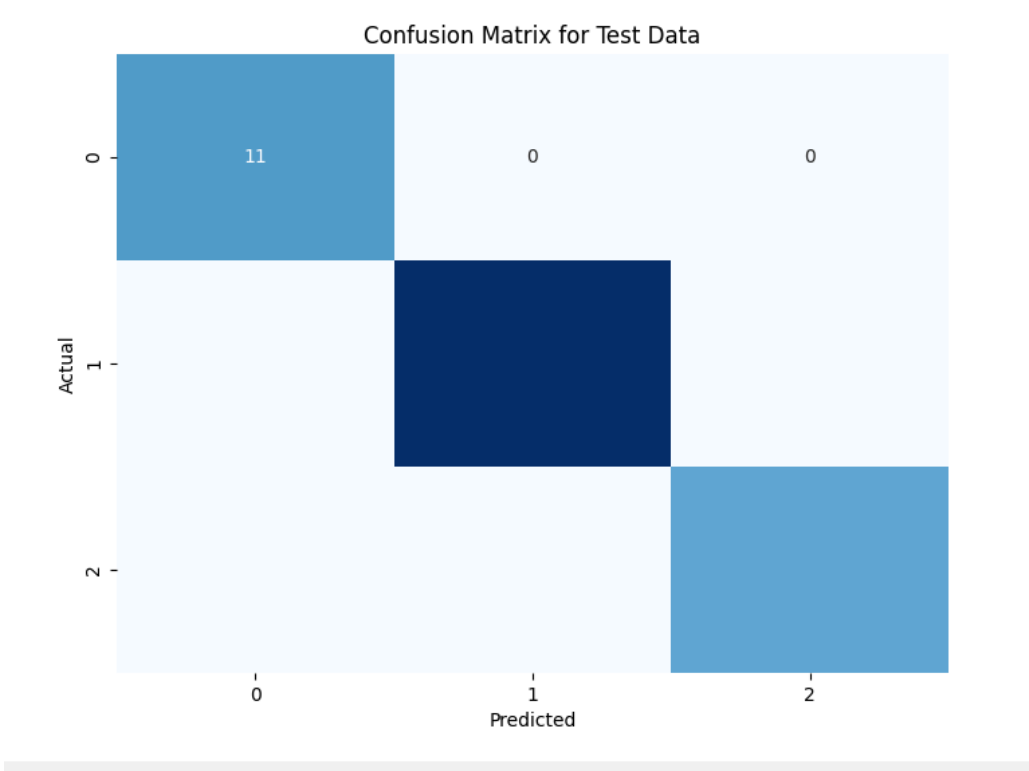
Random Forest Sonuçları:					
	precision	recall	f1-score	support	
0	1.00	1.00	1.00	10	
1	1.00	1.00	1.00	21	
2	1.00	1.00	1.00	9	
accuracy			1.00	40	
macro avg	1.00	1.00	1.00	40	
weighted avg	1.00	1.00	1.00	40	

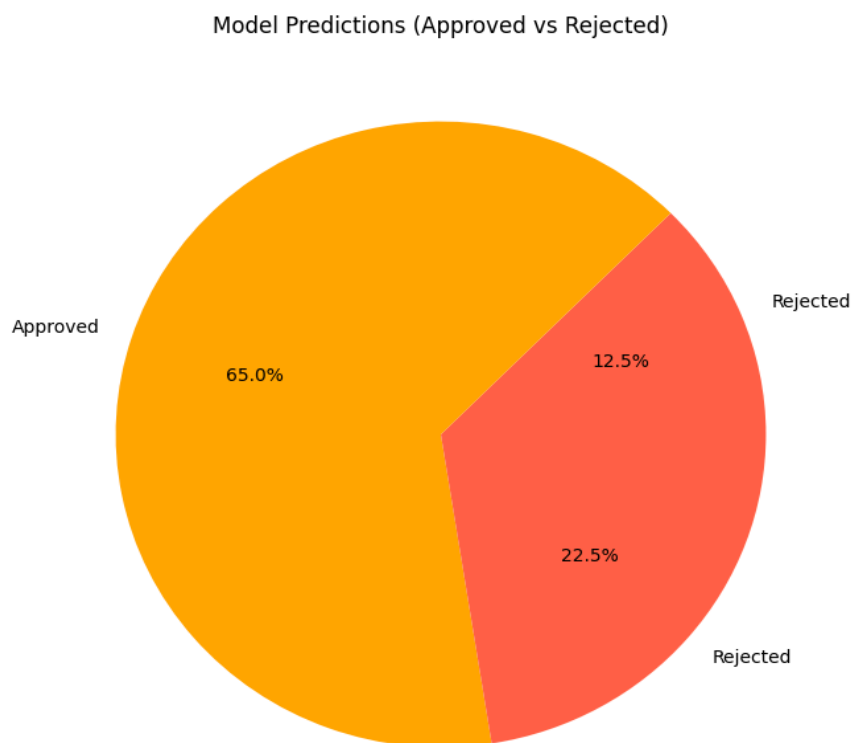
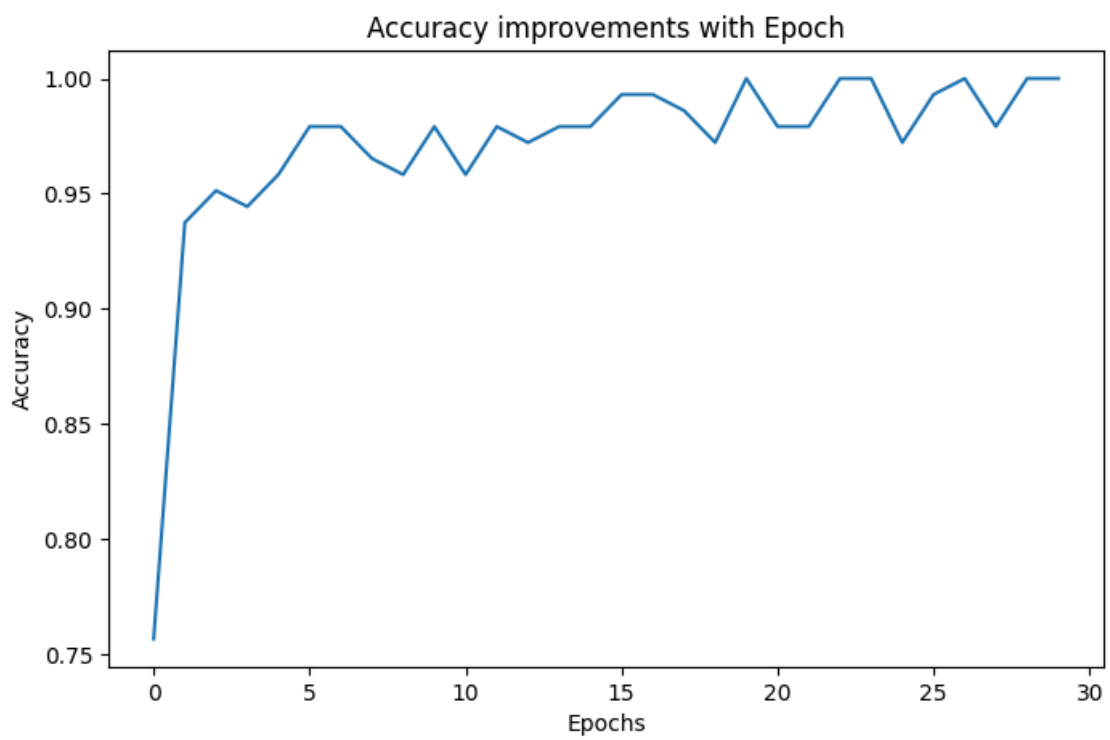
Lojistik Regresyon Sonuçları:

Lojistik Regresyon Sonuçları:					
	precision	recall	f1-score	support	
0	1.00	0.90	0.95	10	
1	0.95	1.00	0.98	21	
2	1.00	1.00	1.00	9	
accuracy			0.97	40	
macro avg	0.98	0.97	0.97	40	
weighted avg	0.98	0.97	0.97	40	

En iyi hiperparametreler: {'C': 1, 'kernel': 'linear'}

Confusion Matrix Sonuçları





Sonuçlar

Bu çalışmada, müşteri segmentasyonunu gerçekleştirmek amacıyla SVM, Karar Ağaçları, Random Forest, Lojistik Regresyon ve Yapay Sinir Ağı algoritmaları kullanılmıştır. Elde edilen sonuçlar, her bir modelin yüksek doğruluk oranlarına ulaştığını göstermektedir. Özellikle, YSA modelinde %95 doğruluk elde edilmiştir. Karar Ağaçları ve Random Forest modelleri de %100 doğruluk oranı ile dikkat çekmiştir. Bu bulgular, müşteri segmentasyonunda kullanılan algoritmaların etkinliğini ve doğruluğunu ortaya koymakta, işletmelerin pazarlama stratejilerini optimize etmelerine katkıda bulunmaktadır.

Kaynaklar

- [1] Kumar, V., & Reinartz, W. (2016). Creating Enduring Customer Value. *Journal of Marketing*, 80(6), 36-68.
- [2] Dhanjal, A. K., & Rao, M. (2019). Customer Segmentation using K-Means and Support Vector Machines. *International Journal of Information Technology*, 11(3), 579-585.
- [3] Kotler, P., & Keller, K. L. (2016). *Marketing Management*. Pearson Education Limited.
- [4] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.