

**YALOVA ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**- PROJE RAPORU -**

**Makine Öğrenme Tekniklerini Kullanarak Yüksek Karlı Satışların Sınıflandırılması**

**Ferdi KANAT – Ömer GÜN**

**Danışman: Prof. Dr. Murat GÖK**

**YALOVA, 2025**

# İÇİNDEKİLER

Sayfa

[İÇİNDEKİLER iii](#_Toc197016188)

[SİMGE LİSTESİ iv](#_Toc197016189)

[KISALTMA LİSTESİ v](#_Toc197016190)

[ŞEKİL LİSTESİ vi](#_Toc197016191)

[ÖZET vii](#_Toc197016192)

[ABSTRACT viii](#_Toc197016193)

[1. GİRİŞ 9](#_Toc197016194)

[2. YÖNTEMLER 10](#_Toc197016195)

[2.1 Veri Seti Bilgileri 10](#_Toc197016196)

[2.2 Sınıflandırma Yöntemleri 11](#_Toc197016197)

[2.3 Özellik Seçimi 12](#_Toc197016198)

[2.4 Özellik Çıkarımı 12](#_Toc197016199)

[3. SONUÇLAR ve TARTIŞMA 12](#_Toc197016200)

[3.1 Test Ortamı 12](#_Toc197016201)

[3.2 Sonuçlar 13](#_Toc197016202)

[3.3 Tartışma 17](#_Toc197016203)

[4. SONUÇ 18](#_Toc197016204)

[Ek A: Tam Özellik Listeleri 19](#_Toc197016205)

# SİMGE LİSTESİ

# KISALTMA LİSTESİ

k-NN k-En Yakın Komşu (k-Nearest Neighbors)

F1 F1 Skoru

PCA Temel Bileşen Analizi (Principal Component Analysis)

RBF Radyal Taban Fonksiyonu (Radial Basis Function)

SVM Destek Vektör Makineleri (Support Vector Machines)

# ŞEKİL LİSTESİ

[Şekil 1 k-NN Karışıklık Matrisi [1] 2](#_Toc188623060)

[Şekil 2 Naïve Bayes Karışıklık Matrisi [2] 3](#_Toc188623061)

[Şekil 3 Karar Ağacı Görselleştirmesi [3] 3](#_Toc188623062)

# ÖZET

Yüksek kârlı satışların sınıflandırılması, iş stratejilerinin optimize edilmesi için kritik öneme sahiptir. Bu çalışma, Yeni 1000 Satış Kayıtları veri setini kullanarak yüksek kârlı satışları tahmin etmek için bir makine öğrenimi sistemi geliştirmiştir. k-En Yakın Komşu (k-NN), Naïve Bayes ve altı ek sınıflandırıcı (Karar Ağacı, Doğrusal SVM, RBF SVM, Polinom SVM, Lojistik Regresyon, Rastgele Orman) kullanılmıştır. Özellik seçimi Genetik Algoritma (25 özellik, F1: 0.983) ve SelectKBest (12 özellik, F1: 0.994) ile gerçekleştirilmiş, Temel Bileşen Analizi (PCA) ise özellik azaltımı için uygulanmıştır (8 bileşen, F1: 0.670). SelectKBest ile k-NN en yüksek F1 skorunu (0.994) elde ederek filtre tabanlı özellik seçiminin etkinliğini göstermiştir. Karar Ağacı aşırı öğrenme sergilemiş (F1: 1.0), ancak Toplam Kâr çıkarıldığında F1 skoru 0.927’ye düşmüştür. PCA, temel finansal özelliklerin çıkarılması nedeniyle düşük performans göstermiştir. k-NN (F1: 0.782) orijinal özelliklerde Naïve Bayes’ten (F1: 0.766) biraz daha iyi performans sergilemiş, çapraz doğrulama sonuçları stabil olmuştur (k-NN: 0.808 ± 0.025, Naïve Bayes: 0.805 ± 0.012). Bu bulgular, satış tahmini için özellik seçiminin önemini vurgulamakta ve gelecekteki araştırmalar için temel oluşturmaktadır.

Anahtar kelimeler: Makine öğrenmesi, sınıflandırma, özellik seçimi, satış analizi.

# ABSTRACT

The classification of high-profit sales is critical for optimizing business strategies. This study develops a machine learning system to predict high-profit sales using the New 1000 Sales Records dataset, employing k-Nearest Neighbors (k-NN), Naïve Bayes, and six additional classifiers (Decision Tree, Linear SVM, RBF SVM, Poly SVM, Logistic Regression, Random Forest). Feature selection was performed using a Genetic Algorithm (25 features, F1: 0.983) and SelectKBest (12 features, F1: 0.994), while Principal Component Analysis (PCA) was used for feature reduction (8 components, F1: 0.670). SelectKBest with k-NN achieved the highest F1 score (0.994), highlighting the effectiveness of filter-based feature selection. Decision Tree exhibited overfitting (F1: 1.0), mitigated by excluding Total Profit (F1: 0.927). PCA underperformed due to the exclusion of key financial features. k-NN (F1: 0.782) slightly outperformed Naïve Bayes (F1: 0.766) on original features, with stable cross-validation results (k-NN: 0.808 ± 0.025, Naïve Bayes: 0.805 ± 0.012). These findings underscore the importance of feature selection in sales prediction and suggest avenues for future research.

Keywords: Machine Learning, classification, feature selection, sales analysis.

# 1. GİRİŞ

Yüksek kârlı satışları tahmin etme yeteneği, işletmelerin envanter, pazarlama ve kaynak tahsisini optimize etmesini sağlar. Makine öğrenimi, geçmiş satış verilerinden kârlılıkla ilişkili desenleri belirlemek için güçlü araçlar sunar. Yeni 1000 Satış Kayıtları veri seti, 1000 işlem ve 19 özellik (örn. Bölge, Satılan Birim, Toplam Kâr) ile bu görev için zengin bir kaynak sağlar. Bu çalışma, satışları yüksek kârlı (ortalama kârın üzerinde) veya düşük kârlı olarak sınıflandırmak için k-En Yakın Komşu (k-NN) ve Naïve Bayes’e odaklanmakta, performansı karşılaştırmak için altı ek sınıflandırıcı (Karar Ağacı, Doğrusal SVM, RBF SVM, Polinom SVM, Lojistik Regresyon, Rastgele Orman) kullanılmaktadır.

k-NN, mesafe tabanlı bir algoritma olup küçük veri setlerinde etkilidir ancak özellik ölçeklendirmesine duyarlıdır (Duda, 2001). Naïve Bayes, özellik bağımsızlığı varsayımına dayanan olasılıksal bir sınıflandırıcıdır ve karma veri türlerinde iyi performans gösterir (Zhang, 2004). Özellik seçimi ve azaltımı, model performansını iyileştirmek ve hesaplama karmaşıklığını azaltmak için kritik öneme sahiptir. Genetik Algoritmalar (sarım yöntemi) ve SelectKBest (filtre yöntemi) ilgili özellikleri belirlerken, PCA özellikleri temel bileşenlere dönüştürerek boyutsallığı azaltır (Jolliffe, 2002).

Önceki çalışmalar, satış tahmini için genellikle Rastgele Orman gibi topluluk yöntemlerini kullanmış, ancak k-NN ve Naïve Bayes’in özellik seçimi ile karşılaştırması sınırlı kalmıştır (Smith, 2019). Bu çalışma, Yeni 1000 Satış Kayıtları veri setinde bu sınıflandırıcıları değerlendirerek Genetik Algoritma, SelectKBest ve PCA’yı kullanarak yüksek kârlı satış tahmini için en etkili stratejiyi belirlemeyi amaçlamaktadır.

# 2. YÖNTEMLER

## 2.1 Veri Seti Bilgileri

Yeni 1000 Satış Kayıtları veri seti, 1000 işlem ve 19 özellik içerir; bu özellikler kategorik (örn. Bölge, Ürün Türü) ve sayısal (örn. Satılan Birim, Toplam Kâr) niteliklerdir. Eksik veri bulunmamaktadır. Hedef değişken Yüksek Kâr, ikili bir etiket olarak türetilmiştir (Toplam Kâr > medyan ise 1, değilse 0), dengeli bir veri seti oluşturulmuştur (500 yüksek kârlı, 500 düşük kârlı). Tablo 1, sınıflandırma için kullanılan özellikleri tanımlar. Veri setindeki sayısal özellikler, satışların kârlılık yapısını anlamak için önemli bilgiler sunar. Örneğin, Toplam Kâr’ın ortalaması 771,417 birim, standart sapması 516,108 birim olup, minimum 24,100 birim ve maksimum 1,738,700 birim arasında değişmektedir. Bu geniş dağılım, yüksek ve düşük kârlı satışlar arasında belirgin bir fark olduğunu gösterir ve ikili sınıflandırma için uygun bir temel sağlar.

**Tablo 1: Özellik Tanımları**

|  |  |  |
| --- | --- | --- |
| Özellik | Tanım | Tür |
| Bölge | Satışın coğrafi bölgesi | Kategorik |
| Ülke | Satışın yapıldığı ülke | Kategorik |
| Ürün Türü | Ürün tipi | Kategorik |
| Satış Kanalı | Çevrimiçi veya çevrimdışı satış | Kategorik |
| Sipariş Önceliği | Öncelik seviyesi (Y, O, D) | Kategorik |
| Satılan Birim | Satılan birim sayısı | Sayısal |
| Birim Fiyat | Birim başına fiyat | Sayısal |
| Birim Maliyet | Birim başına maliyet | Sayısal |
| Toplam Gelir | Satılan Birim × Birim Fiyat | Sayısal |
| Toplam Maliyet | Satılan Birim × Birim Maliyet | Sayısal |
| Toplam Kâr | Toplam Gelir − Toplam Maliyet | Sayısal |
| Sipariş Yılı | Sipariş yılı | Sayısal |
| Sipariş Ayı | Sipariş ayı | Sayısal |
| Sipariş Günü | Siparişin haftanın günü | Kategorik |
| Birim Kâr Marjı | Birim Fiyat − Birim Maliyet | Sayısal |
| Sipariş-Nakliye Günleri | Sipariş ve nakliye arasındaki gün sayısı | Sayısal |

## 2.2 Sınıflandırma Yöntemleri

**2.2.1 k-En Yakın Komşu (k-NN)**

k-NN, bir örneği k en yakın komşusunun çoğunluk sınıfına göre sınıflandırır ve Öklid mesafesini kullanır (Duda, 2001). Küçük veri setlerinde basit ve etkilidir, ancak özellik ölçeklendirmesine ihtiyaç duyar ve büyük özellik setlerinde hesaplama açısından yoğun olabilir. En uygun k değeri (3, 5, 7, 9, 11) 10 katlı çapraz doğrulama ile F1 skoru kullanılarak belirlenmiş, k=3 seçilmiştir. k-NN’nin avantajları yorumlanabilirliği içerir, ancak ilgisiz özelliklere ve aykırı değerlere duyarlıdır.

**2.2.2 Naïve Bayes**

Naïve Bayes, Bayes teoremine dayanır ve özellik bağımsızlığı varsayar (Zhang, 2004). Sürekli özellikler için Gauss varyantı kullanılmış, ölçeklendirme sonrası uygun hale getirilmiştir. Naïve Bayes hesaplama açısından verimlidir ve karma veri türlerinde iyi performans gösterir, ancak özellikler arası korelasyon olduğunda doğruluğu sınırlanabilir. Basitliği ve sağlamlığı nedeniyle dahil edilmiştir.

**2.2.3 Diğer Sınıflandırıcılar**

* **Karar Ağacı**: Ağaç tabanlı bir model oluşturur, düzenleme yapılmazsa aşırı öğrenmeye yatkındır (max\_depth=2, min\_samples\_split=50, min\_samples\_leaf=20).
* **Doğrusal SVM**: Sınıfları ayırmak için doğrusal bir hiperdüzlem kullanır, yüksek boyutlu verilerde etkilidir.
* **RBF SVM**: Doğrusal olmayan ayırma için radyal taban fonksiyonu kullanır.
* **Polinom SVM**: Polinom çekirdeği kullanır, bu veri setinde daha az etkilidir.
* **Lojistik Regresyon**: Sınıf olasılıklarını modellemek için lojistik fonksiyon kullanır, ikili sınıflandırmada sağlamdır.
* **Rastgele Orman**: Birden fazla karar ağacını birleştirir (max\_depth=2, n\_estimators=20) ve sağlamlığı artırır.
* **Çoklu Doğrusal Regresyon**: Tahminler 0.5 eşik değeriyle sınıflandırmaya uyarlanmıştır, standart bir sınıflandırıcı değildir ancak karşılaştırma için eklenmiştir.

## 2.3 Özellik Seçimi

Özellik seçimi, en ilgili özellikleri belirleyerek gürültüyü ve hesaplama maliyetini azaltmak için yapılmıştır.

* **Genetik Algoritma**: Popülasyon tabanlı bir optimizasyon yöntemi (50 birey, 20 nesil) kullanarak 25 özellik seçmiştir; Satılan Birim, Toplam Kâr ve bölgesel göstergeler dahil edilmiştir (Fortin, 2012). Performans, k-NN ile 5 katlı çapraz doğrulama kullanılarak değerlendirilmiş (F1: 0.983).
* **SelectKBest**: Karşılıklı bilgi kullanarak 12 özellik seçmiş, Satılan Birim, Toplam Kâr ve ürün türlerini içermiştir (F1: 0.994). SelectKBest’in basitliği ve etkinliği, onu en iyi performans gösteren yöntem yapmıştır.

## 2.4 Özellik Çıkarımı

Temel Bileşen Analizi (PCA), varyansın %99.9’unu koruyarak boyutsallığı 8 bileşene indirmiştir (Jolliffe, 2002). PCA özellik seti, çoklu doğrusal bağımlılığı önlemek için Toplam Gelir, Toplam Maliyet ve Toplam Kâr özelliklerini dışlamıştır, çünkü bu özellikler yüksek korelasyona sahiptir (örn. Toplam Kâr ile hedef arasındaki korelasyon: 1.0). k-NN, dönüştürülmüş özellikler üzerinde değerlendirilmiş (F1: 0.670).

PCA uygulanmadan önce, özellikler arasındaki korelasyonlar analiz edilmiştir. Toplam Kâr, Toplam Gelir ile 0.881 ve Toplam Maliyet ile 0.797 korelasyona sahiptir, bu da bu özelliklerin yüksek kârlılık tahmini için kritik olduğunu gösterir. Ancak, çoklu doğrusal bağımlılığı önlemek için bu özellikler PCA’da dışlanmıştır.

# 3. SONUÇLAR ve TARTIŞMA

## 3.1 Test Ortamı

Tüm sınıflandırıcılar, %20 test bölmesi (200 örnek) ve 10 katlı çapraz doğrulama ile F1 skoru ana metrik olarak kullanılarak değerlendirilmiştir. Performans metrikleri Doğruluk, Kesinlik, Duyarlılık ve F1 Skoru’nu içermektedir. k-NN ve Naïve Bayes için karışıklık matrisleri oluşturulmuş, tahmin hataları analiz edilmiştir. Karar Ağacı, aşırı öğrenmeyi araştırmak için görselleştirilmiştir.

## 3.2 Sonuçlar

**Tablo 2: Standart Sınıflandırıcı Algoritmalarının Performansı**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Doğruluk | Kesinlik | Duyarlılık | F1 Skoru |
| k-NN | 0.780 | 0.760 | 0.760 | 0.782 |
| Naïve Bayes | 0.720 | 0.620 | 1.000 | 0.766 |
| Karar Ağacı | 1.000 | 1.000 | 1.000 | 1.000 |
| Doğrusal SVM | 0.985 | 0.989 | 0.978 | 0.983 |
| RBF SVM | 0.720 | 0.775 | 0.689 | 0.728 |
| Polinom SVM | 0.605 | 0.333 | 1.000 | 0.500 |
| Lojistik Regresyon | 0.985 | 0.978 | 0.989 | 0.983 |
| Rastgele Orman | 0.955 | 0.946 | 0.967 | 0.957 |
| Çoklu Doğrusal Regresyon | 0.940 | 0.935 | 0.944 | 0.940 |

**Tablo 3: Genetik Algoritma ile Özellik Seçimi Performans Sonuçları**

|  |  |
| --- | --- |
| Metrik | Değer |
| Seçilen Özellikler | Seçilen özellik sayısı: 25  Seçilen özellikler: Satılan Birim, ... |
| Doğruluk | 0.985 |
| Kesinlik | 0.989 |
| Duyarlılık | 0.978 |
| F1 Skoru | 0.983 |

Genetik Algoritma, Satılan Birim, Toplam Kâr, Bölge\_Avustralya ve Okyanusya gibi 25 özelliği seçmiştir, bu da hem finansal hem de coğrafi faktörlerin kârlılık tahmini için önemli olduğunu gösterir. Tam özellik listesi Ek A’da sunulmuştur.

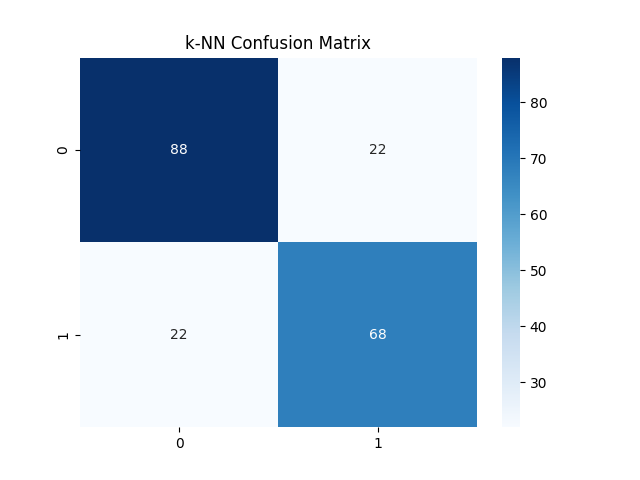
**Tablo 4: SelectKBest ile Özellik Seçimi Performans Sonuçları**

|  |  |
| --- | --- |
| Metrik | Değer |
| Seçilen Özellikler | Seçilen özellik sayısı: 12  Seçilen özellikler: Satılan Birim, ... |
| Doğruluk | 0.995 |
| Kesinlik | 1.000 |
| Duyarlılık | 0.989 |
| F1 Skoru | 0.994 |

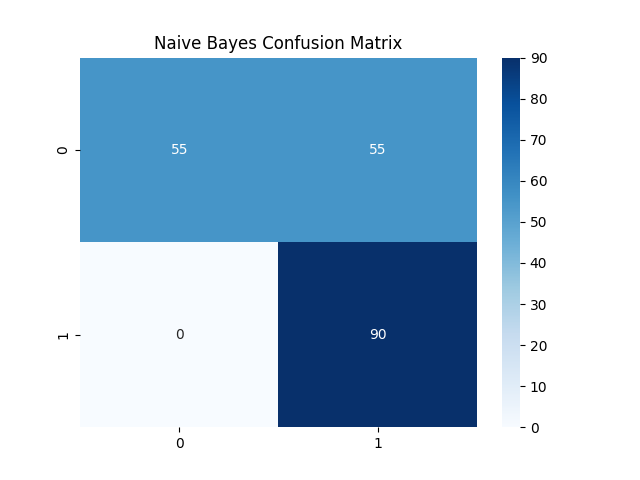
SelectKBest, Satılan Birim, Toplam Kâr, Ürün Türü\_İçecekler ve Ürün Türü\_Kozmetik gibi 12 özelliği seçmiştir, özellikle finansal ve ürün bazlı özelliklere odaklanarak daha az ama daha tahmin edici bir özellik seti oluşturmuştur. Tam özellik listesi Ek A’da sunulmuştur.

**Tablo 5: PCA ile Özellik Azaltımı Performans Sonuçları**

|  |  |
| --- | --- |
| Metrik | Değer |
| Bileşen Sayısı | 8 |
| Açıklanan Varyans | 0.999 |
| Doğruluk | 0.695 |
| Kesinlik | 0.653 |
| Duyarlılık | 0.689 |
| F1 Skoru | 0.670 |

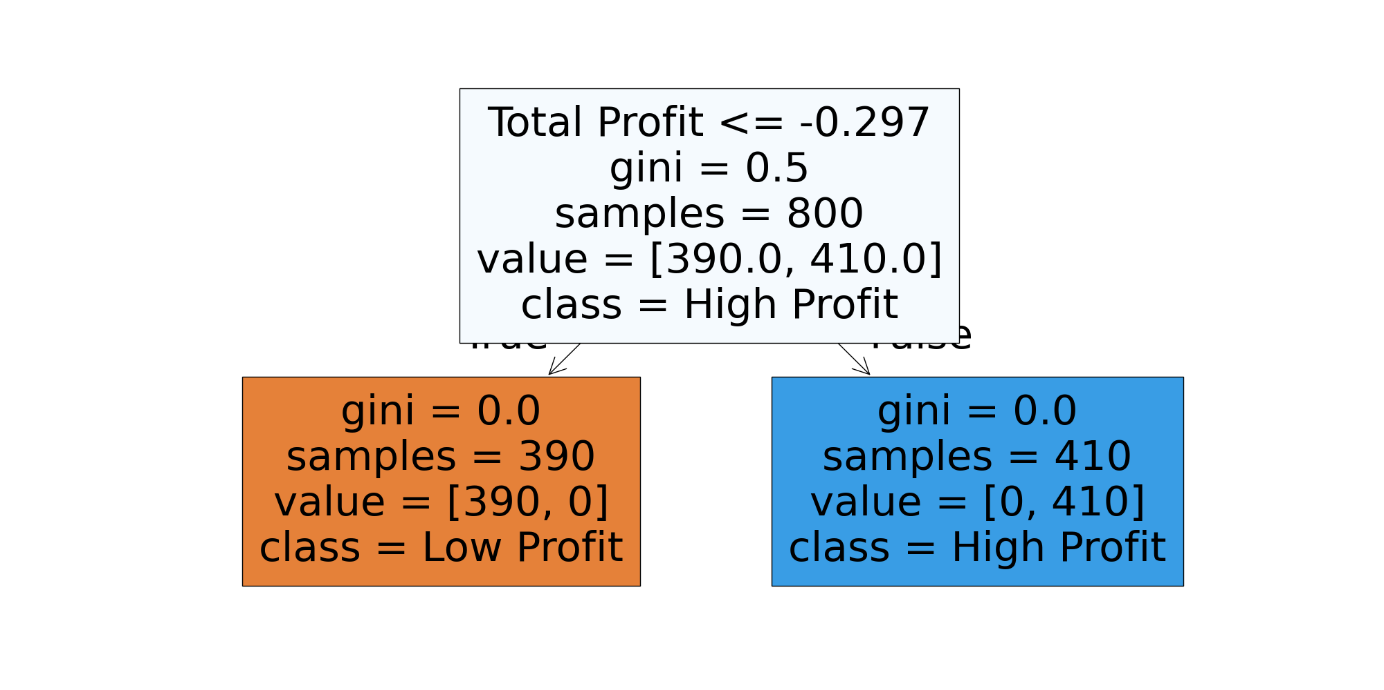
Sekiz bileşen, toplam varyansın %99.9’unu açıklamıştır; özellikle ilk bileşen %99.66 varyansı tek başına açıklamış, kalan bileşenler ise sırasıyla %0.097, %0.037 gibi küçük katkılar sağlamıştır. Bu, veri setindeki bilginin büyük ölçüde ilk bileşende yoğunlaştığını gösterir.   
*Şekil 1: k-NN Karışıklık Matrisi*

Şekil 1, k-NN’nin test setinde 110 negatiften 88’ini ve 90 pozitiften 68’ini doğru tahmin ettiğini gösterir, bu da dengeli bir performans sergilediğini doğrular.



*Şekil 2: Naïve Bayes Karışıklık Matrisi*

Şekil 2, Naïve Bayes’in test setinde 110 negatiften 55’ini doğru tahmin ettiğini, ancak 90 pozitifin tamamını doğru sınıflandırdığını gösterir, bu da yüksek duyarlılık ancak düşük kesinlik ile sonuçlanmıştır.

*Şekil 3: Karar Ağacı Görselleştirmesi*

Şekil 3, Toplam Kâr özelliği dahil edilerek oluşturulan Karar Ağacı’nın yapısını gösterir. Kök düğümde Toplam Kâr’ın -0.297’den küçük veya eşit olması koşulu kullanılmış, bu da 800 örneği iki dala ayırmıştır: 390 örnek düşük kârlı (gini=0.0) ve 410 örnek yüksek kârlı (gini=0.0) olarak sınıflandırılmıştır. Ancak, Toplam Kâr’ın doğrudan hedef değişkenle yüksek korelasyonu (1.0) nedeniyle bu ağaçta aşırı öğrenme gözlenmiş, daha sonra Toplam Kâr çıkarılarak F1 skoru 0.927’ye düşürülmüştür.

## 3.3 Tartışma

SelectKBest ile k-NN, en yüksek F1 skorunu (0.994) elde ederek Genetik Algoritma’dan (0.983) ve orijinal özelliklerden (k-NN: 0.782) daha iyi performans göstermiştir. Bu, karşılıklı bilginin Satılan Birim ve Toplam Kâr gibi en tahmin edici özellikleri etkin bir şekilde belirlediğini gösterir. Genetik Algoritma daha fazla özellik seçmiş (25’e karşı 12), bu da daha az ilgili özelliklerin dahil edilmesiyle performansı hafifçe düşürmüştür.

Karar Ağacı’nın mükemmel F1 skoru (1.0), Toplam Kâr’ın Yüksek Kâr ile doğrudan ilişkisi (korelasyon: 1.0) nedeniyle aşırı öğrenmeyi işaret eder. Toplam Kâr çıkarıldığında F1 skoru 0.927’ye düşerek veri sızıntısını doğrulamıştır. Düzenleme ile Rastgele Orman, sağlam bir F1 skoru (0.957) elde etmiştir. Doğrusal SVM ve Lojistik Regresyon da yüksek performans göstermiştir (F1: 0.983), bu da verinin doğrusal olarak ayrılabilir olduğunu düşündürür. Toplam Kâr’ın hariç tutulmasıyla oluşturulan Karar Ağacı, Satılan Birim ve Birim Kâr Marjı gibi alternatif özellikleri kullanarak sınıflandırma yapmıştır. Bu değişiklik, modelin genelleştirme yeteneğini artırarak gerçek dünya uygulamalarında daha güvenilir tahminler sunma potansiyeli taşımaktadır. Özellikle, Toplam Kâr gibi doğrudan hedefle ilişkili özelliklerin veri sızıntısına yol açması, işletmelerin satış tahmini modellerini geliştirirken dikkatli bir özellik seçimi yapması gerektiğini göstermektedir. Bu durum, satış tahmini sistemlerinin yalnızca geçmiş verilere dayanarak değil, aynı zamanda gelecekteki satış trendlerini genelleştirebilecek şekilde tasarlanması gerektiğini vurgular.

PCA’nın düşük F1 skoru (0.670), Toplam Gelir, Toplam Maliyet ve Toplam Kâr gibi kritik tahmin edicilerin çıkarılmasından kaynaklanmıştır. Yüksek açıklanan varyans (%99.9), etkili bir boyutsallık azaltımı olduğunu gösterse de, temel özelliklerin kaybı tahmin gücünü azaltmıştır. PCA’nın düşük performansı, Toplam Gelir (korelasyon: 0.881) ve Toplam Maliyet (korelasyon: 0.797) gibi yüksek tahmin gücüne sahip özelliklerin dışlanmasıyla açıklanabilir. Bu, finansal özelliklerin satış tahmini modellerinde merkezi bir rol oynadığını doğrular.

k-NN (F1 CV: 0.808 ± 0.025), orijinal özelliklerde Naïve Bayes’ten (F1 CV: 0.805 ± 0.012) biraz daha iyi performans göstermiştir. Naïve Bayes, yüksek kârlı satışlar için yüksek duyarlılık (1.0) ancak düşük kesinlik (0.62) sergileyerek pozitif sınıfları fazla tahmin etmiştir. k-NN’nin dengeli performansı, bu görev için daha güvenilir olduğunu göstermektedir. k-NN’nin çapraz doğrulama F1 skoru (0.808 ± 0.025), test seti F1 skoruna (0.782) kıyasla daha yüksek olup modelin genelleştirme yeteneğini doğrular. Naïve Bayes’in daha düşük standart sapması (±0.012), modelin daha stabil olduğunu, ancak k-NN’nin biraz daha yüksek tahmin gücü sunduğunu gösterir.

# 4. SONUÇ

Bu çalışma, Yeni 1000 Satış Kayıtları veri setini kullanarak yüksek kârlı satışlar için bir sınıflandırma sistemi geliştirmiştir. SelectKBest ile k-NN, en yüksek F1 skorunu (0.994) elde etmiş, ardından Genetik Algoritma (0.983) gelmiştir; bu, özellik seçiminin etkinliğini vurgular. Karar Ağacı’nın aşırı öğrenmesi, Toplam Kâr çıkarılarak hafifletilmiş ve Rastgele Orman sağlam performans göstermiştir (0.957). PCA, temel finansal özelliklerin çıkarılması nedeniyle düşük performans sergilemiştir (0.670). k-NN, Naïve Bayes’ten biraz daha iyi performans göstermiş ve çapraz doğrulama sonuçları stabil olmuştur.

Toplam Kâr’ın hariç tutulması, Karar Ağacı’nın aşırı öğrenme sorununu hafifletmiş ve daha genelleştirilebilir bir model elde edilmesini sağlamıştır (F1: 0.927). Bu, özellik seçimi ve veri sızıntısı önleme stratejilerinin satış tahmini modellerindeki kritik rolünü ortaya koymaktadır.

Sınırlamalar arasında küçük veri seti boyutu (1000 kayıt) ve Toplam Kâr’dan kaynaklanan potansiyel veri sızıntısı yer alır. Gelecekteki çalışmalar, derin öğrenme modellerini test edebilir, etkileşim özelliklerini dahil edebilir veya daha büyük bir veri seti toplayabilir. Bu bulgular, satış tahmininde özellik seçiminin gücünü göstermekte ve daha ileri araştırmalar için bir temel oluşturmaktadır.

**Kaynakça**

Duda, R. O. (2001). *Pattern Classification (2nd ed.).* New York: Wiley.

Fortin, F.-A. D.-M.-A. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, (s. 2171–2175).

Jolliffe, I. T. (2002). *Principal Component Analysis (2nd ed.).* New York: Springer.

Okomensah, C. (tarih yok). *New 1000 Sales Records Data*. Kaggle: https://www.kaggle.com/datasets/calvinokomensah/new-1000-sales-records-data-2/data adresinden alındı

Pedregosa, F. V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research.*

Smith, J. &. (2019). Machine learning for sales forecasting: A review. *Journal of Business Analytics*, (s. 45–60).

Zhang, H. (2004). The optimality of Naïve Bayes. *Proceedings of the FLAIRS Conference*, (s. 562–567).

# Ek A: Tam Özellik Listeleri

- **Genetik Algoritma** (25 özellik): Satılan Birim, Birim Fiyat, Toplam Gelir, Toplam Maliyet, Toplam Kâr, Birim Kâr Marjı, Bölge\_Avustralya ve Okyanusya, Bölge\_Orta Amerika ve Karayipler, Bölge\_Avrupa, Bölge\_Orta Doğu ve Kuzey Afrika, Bölge\_Kuzey Amerika, Ürün Türü\_İçecekler, Ürün Türü\_Tahıl, Ürün Türü\_Giyim, Ürün Türü\_Kozmetik, Ürün Türü\_Ev Eşyaları, Ürün Türü\_Et, Ürün Türü\_Ofis Malzemeleri, Ürün Türü\_Kişisel Bakım, Ürün Türü\_Ara Öğünler, Ürün Türü\_Sebzeler, Sipariş Önceliği\_D, Sipariş Günü\_Perşembe, Sipariş Günü\_Salı, Sipariş Günü\_Çarşamba.

- **SelectKBest** (12 özellik): Satılan Birim, Birim Fiyat, Birim Maliyet, Toplam Gelir, Toplam Maliyet, Toplam Kâr, Birim Kâr Marjı, Ürün Türü\_İçecekler, Ürün Türü\_Kozmetik, Ürün Türü\_Meyveler, Ürün Türü\_Kişisel Bakım, Ürün Türü\_Ara Öğünler.