



**T.C. BURSA TEKNİK ÜNİVERSİTESİ  
MÜHENDİSLİK VE DOĞA BİLİMLERİ  
FAKÜLTESİ**

**BLM0463 Veri Madenciliğine Giriş Dersi – Dönem Projesi  
Karar Ağacı Sınıflandırma Yöntemi ile Bankacılık Verisi  
Üzerinden Tahminleme**

**Hazırlayan:** Simge SÖYLER

**Öğrenci No:** 21392380026

**Danışman:** Doç. Dr. Erdem YAVUZ

## İÇİNDEKİLER

### Sayfa

ŞEKİLLER DİZİNİ .....	1
1. GİRİŞ .....	1
2. VERİ SETİ TANITIMI .....	3
3. KULLANILAN YÖNTEM: KARAR AĞACI VE ORANGE PLATFORMU .....	5
4. MODEL PERFORMANSI VE DEĞERLENDİRME .....	8
5. SONUÇ VE ÖNERİLER .....	9
6. LİTERATÜR KARŞILAŞTIRMASI .....	10
7. KAYNAKÇA .....	12

## ŞEKİL DİZİNİ

### Sayfa

Şekil 3.1. : Orange platformunda kullanılan karar ağacı modeline ait akış diyagramı.	5
Şekil 3.2. : Orange Select Columns Modülünde, Veri Setindeki Nitelikler İçinden Hedef Değişkenin (y) Belirlenmesi ve Kalan Özniteliklerin Bağımsız Değişken (features) Olarak Atanması.....	5
Şekil 3.3. Karar Ağacı Modeli İçin Belirlenen Parametre Ayarları (Yaprak sayısı, maksimum derinlik, bölünme sınırları) .....	6
Şekil 4.4. : Karar Ağacı Modeline Ait Test Sonuçları ve Metrik Değerlendirme.....	8

## 1. GİRİŞ

Dijital dönüşümle birlikte, kurumların rekabet avantajı elde edebilmesi ve stratejik kararlar alabilmesi için veri odaklı yaklaşımlar benimsemeleri hayati bir gereklilik haline gelmiştir. Artan dijitalleşme, kurumlara her geçen gün daha fazla veri oluşturma, toplama ve bu verileri analiz ederek anlamlı bilgilere dönüştürme fırsatı sunmaktadır. Bu noktada veri madenciliği, farklı kaynaklardan elde edilen büyük veri kümelerinden önceden bilinmeyen ama potansiyel olarak faydalı bilgilerin ortaya çıkarılmasını sağlayan bir disiplindir.

Veri madenciliği, özellikle iş zekası, pazarlama stratejileri, finansal tahminleme, müşteri ilişkileri yönetimi gibi alanlarda yaygın olarak kullanılmakta olup, şirketlere sadece geçmişini analiz etme değil, aynı zamanda geleceği öngörme becerisi de kazandırmaktadır. Bu teknik sayesinde kurumlar; hangi ürünlerin daha çok tercih edildiğini, müşterilerin hangi kampanyalara daha duyarlı olduğunu, riskli kredi başvurularını ya da dolandırıcılık faaliyetlerini öngörebilmekte, bu bilgiler ışığında kaynaklarını daha etkin yönetebilmektedir.

Bankacılık sektörü de bu veri odaklı dönüşümden önemli ölçüde etkilenmiştir. Geleneksel bankacılık anlayışından çıkarak müşteri odaklı hizmet sunumu benimseyen bankalar, büyük veri analitiği, yapay zeka ve makine öğrenmesi gibi teknolojilerle karar süreçlerini zenginleştirmektedir. Özellikle telepazarlama kampanyalarında, hangi müşterilerin yeni bir ürün ya da hizmete ilgi göstereceğini önceden bilmek hem maliyetleri azaltmakta hem de müşteri memnuniyetini artırmaktadır. Bu nedenle, bankalar sahip oldukları veri birikimini doğru analiz ederek stratejik kararlar almak için veri madenciliği tekniklerinden faydalanmaktadır.

Bu proje kapsamında, Portekiz'e ait gerçek bankacılık verilerinden oluşan ve UCI Machine Learning Repository platformunda yayınlanan "Bank Marketing" veri seti kullanılmıştır. Veri seti, belirli bir kampanya kapsamında müşterilere telefon yoluyla erişilip, vadeli mevduat hesabı açılması konusunda ikna edilmeye çalışıldığı görüşmeleri kapsamaktadır. Amaç, her bir müşteri kaydı için, kampanyaya olumlu yanıt verip vermeyeceğini ("yes" ya da "no") tahmin etmektir. Veri seti bu anlamda sınıflandırma problemlerine örnek teşkil etmekte ve farklı algoritmalarla çözüm geliştirilebilecek zengin bir yapıya sahiptir.

Veri setinde yer alan öznitelikler; yaş, meslek, medeni durum, eğitim durumu gibi demografik bilgilerden, kampanya detaylarına (arama türü, görüşme günü, ayı, süresi, önceki kampanyalara katılım durumu vb.) kadar çeşitli bilgileri içermektedir. Bu bilgilerin analiz edilmesiyle, hangi özniteliklerin kampanyaya "evet" deme kararında daha etkili olduğu anlaşılabilmektedir. Bu proje kapsamında, bu özniteliklerin etkisini görebilmek adına karar ağacı algoritması kullanılmıştır.

Karar ağacı algoritması, her bir özniteliği dikkate alarak veriyi dallara ayırır ve en sonunda müşterinin kampanyaya olumlu yanıt verip vermeyeceğini tahmin etmeye çalışır. Bu modelin en önemli avantajlarından biri, elde edilen sonuçların kolay yorumlanabilir ve görselleştirilebilir olmasıdır.

Bu sayede sadece teknik uzmanlar değil, pazarlama ya da yönetim ekipleri de modelin mantığını anlayarak stratejik kararlar alabilir.

Bu projede karar ağacı algoritmasının uygulanabilirliği Orange veri madenciliği platformu aracılığıyla gerçekleştirilmiştir. Orange, görsel tabanlı, sürükle-bırak mantığıyla çalışan, akademik ve endüstriyel kullanıma uygun, Python tabanlı bir veri madenciliği aracıdır. Özellikle makine öğrenmesi, sınıflandırma, kümeleme ve görselleştirme gibi yöntemlerin uygulanmasını kolaylaştıran arayüzü sayesinde, veri bilimi alanına yeni başlayanlar için öğretici ve etkili bir platformdur.

Bu çalışma ayrıca Elif Akkaya ve Safiye Turgay tarafından WSEAS Transactions on Computers dergisinde yayınlanan "Unveiling the Power: A Comparative Analysis of Data Mining Tools through Decision Tree Classification on the Bank Marketing Dataset" başlıklı akademik makaleden ilham almıştır. Söz konusu makalede, Orange platformu başta olmak üzere, Weka, KNIME ve RapidMiner gibi farklı veri madenciliği yazılımları karşılaştırılmakta, karar ağacı algoritması üzerindeki performansları değerlendirilmektedir. Yazarlar, bu platformların hem model başarımlarını hem de kullanıcı dostu arayüzlerini analiz etmiş, Orange'ın sezgisel arayüzü sayesinde öğrenme eğrisinin daha hızlı olduğuna dikkat çekmiştir. Bu nedenle, bu proje çalışması için Orange platformunun seçilmesi hem akademik hem de uygulamalı açıdan rasyonel bir tercihtir.

## **2. VERİ SETİ TANITIMI**

Bu çalışmada kullanılan veri seti, Portekiz'deki bir bankanın 2008-2010 yılları arasında gerçekleştirdiği telepazarlama kampanyalarına ait verilerden oluşmaktadır.

Veri seti, University of California Irvine (UCI) Machine Learning Repository'de yayınlanmış olup, akademik çalışmalarda sıklıkla kullanılan, denge problemi gösteren, çok boyutlu ve hem sayısal hem de kategorik veriler içeren bir sınıflandırma problemidir.

Veri setine [su linkten](#) erişilebilmektedir:



bank+marketing (1).zip

Veri seti toplam **45.211** gözlem (satır) ve **17** öznitelikten (sütun) oluşmaktadır. Bu öznitelikler, müşterilere ait demografik bilgilerden ("age", "job", "marital", "education") finansal bilgilere ("balance", "housing", "loan"), kampanya özelliklerinden ("contact", "month", "day", "duration", "campaign", "pdays", "previous") önceki kampanya başarılarına kadar (örn: "poutcome") geniş bir çerçevede dağılmaktadır. Bu durum, veri setini hem veri ön işleme açısından hem de sınıflandırma performanslarını gözlemleme bakımından öğretici bir kaynak haline getirmektedir.

Veri setinde yer alan **hedef değişken "y"** sütunu, her müşteri için kampanya sonucunu belirtmektedir. Bu sütun iki değer alabilmektedir: **"yes"** (müşteri kampanyayı kabul etti) veya **"no"** (müşteri kampanyayı reddetti). Bu yönüyle ikili (binary) sınıflandırma problemi niteliği taşımaktadır. Veri setindeki bu dağılım dengesizdir; "no" sınıfının oranı yaklaşık %88 civarında olup, "yes" sınıfı sadece %12 civarında gözlemlenmektedir. Bu nedenle, modelleme süreçlerinde **sınıf dengesizliği** sorunu dikkate alınmalı ve ölçütler buna göre değerlendirilmelidir.

Veri setinde eksik değerler doğrudan boş bırakılmak yerine, kategorik "unknown" etiketi ile işaretlenmiştir. Bu, geleneksel eksik veri tespiti yöntemlerinden farklı olarak, bilinmeyen değerlerin de potansiyel bir bilgi taşıyabileceği varsayımıyla hareket etmeyi gerektirir. "job", "education", "contact" ve "poutcome" sütunlarında bu türden "unknown" değerlere rastlanmakta olup, bu değerlerin modele dahil edilip edilmemesi ön işleme sürecinde veriye göre karar verilmiştir. Orange platformunda bu değerler genellikle kategorik sınıflandırma yapıldığı için ayrı bir kategori olarak kabul edilmiştir.

Nitekim, Elif Akkaya ve Safiye Turgay tarafından yayınlanan ve bu projeye örnek teşkil eden "Unveiling the Power: A Comparative Analysis of Data Mining Tools through Decision Tree Classification on the Bank Marketing Dataset" başlıklı makalede de aynı veri seti kullanılmıştır.

Söz konusu çalışmada, Orange başta olmak üzere farklı veri madenciliği platformları karar ağacı algoritması ile karşılaştırılmış, öğrenme süreci, kullanıcı deneyimi ve performans kriterleri açısından değerlendirme yapılmıştır.

Bu özniteliklerden bazıları:

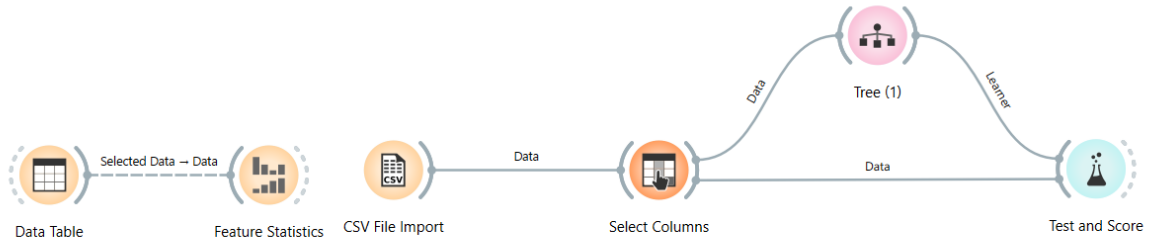
- **age** (yaş),
- **job** (meslek),
- **marital** (medeni durum),
- **education** (eğitim düzeyi),
- **balance** (hesap bakiyesi),
- **housing** (konut kredisi varlığı),
- **loan** (kişisel kredi varlığı) ve
- **contact** (iletişim türü) gibi nitel ve nicel değişkenlerden oluşmaktadır. Bu öznelitlikler kampanyaya verilecek yanıtı etkileyebilecek potansiyel belirleyicilerdir ve karar ağacı modeli için girdi olarak kullanılmıştır.

Veri setinde yer alan **hedef değişken "y"** sütunu, her müşteri için kampanya sonucunu belirtmektedir. Bu sütun iki değer alabilmektedir: **"yes"** (müşteri kampanyayı kabul etti) veya **"no"** (müşteri kampanyayı reddetti). Bu yönüyle ikili (binary) sınıflandırma problemi niteliği taşımaktadır. Veri setindeki bu dağılım dengesizdir; "no" sınıfının oranı yaklaşık %88 civarında olup, "yes" sınıfı sadece %12 civarında gözlemlenmektedir. Bu nedenle, modelleme süreçlerinde **sınıf dengesizliği** sorunu dikkate alınmalı ve ölçütler buna göre değerlendirilmelidir.

Veri setinde eksik değerler doğrudan boş bırakılmak yerine, kategorik **"unknown"** etiketi ile işaretlenmiştir. Bu, geleneksel eksik veri tespiti yöntemlerinden farklı olarak, bilinmeyen değerlerin de potansiyel bir bilgi taşıyabileceği varsayımıyla hareket etmeyi gerektirir. "job", "education", "contact" ve "outcome" sütunlarında bu türden "unknown" değerlere rastlanmakta olup, bu değerlerin modele dahil edilip edilmemesi ön işleme sürecinde veriye göre karar verilmiştir. Orange platformunda bu değerler genellikle kategorik sınıflandırma yapıldığı için ayrı bir kategori olarak kabul edilmiştir. Modellemeden önce bu alanlar dikkatle analiz edilmiş, eksik veri probleminin model performansını etkilememesi için gerekli temizlik ve dönüşümler sağlanmıştır.

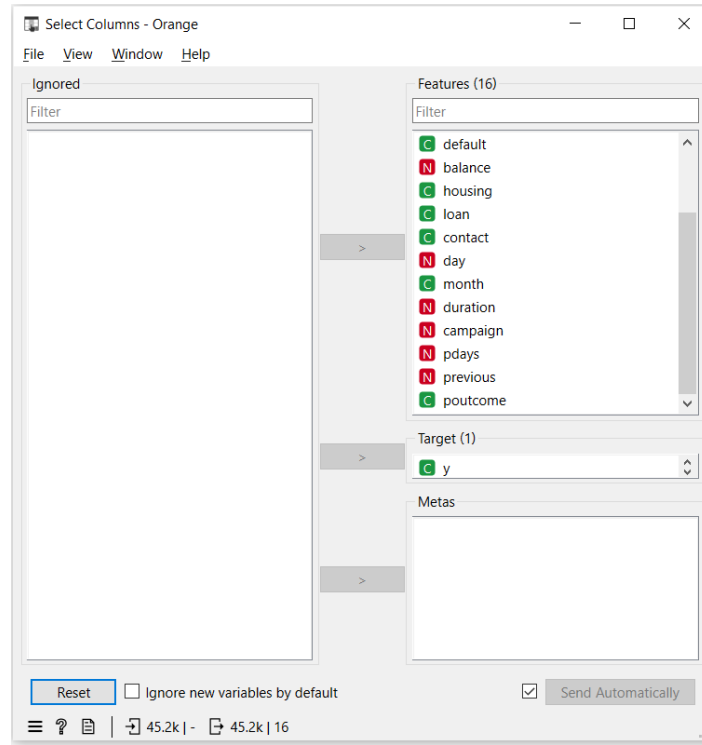
Veri seti, yalnızca veri madenciliği tekniklerinin uygulanmasına değil; aynı zamanda model performanslarının, sınıf dengesizliklerinin, karar ağacı algoritmalarının yorumlanabilirliğinin ve görselleştirmenin etkili şekilde analiz edilebilmesine olanak sağlamaktadır. Bu nedenle, akademik literatürde de bu veri setine sıkça yer verilmektedir.

### 3. KULLANILAN YÖNTEM: KARAR AĞACI VE ORANGE PLATFORMU



**Şekil 3.1.** Orange platformunda kullanılan karar ağacı modeline ait akış diyagramı.

Bu şekil, Orange platformunda oluşturulan karar ağacı modelinin genel akışını göstermektedir. CSV dosyası ile veri içeri alınır, ardından öznelik seçimi (Select Columns) yapılır. Modelleme adımı olarak “Tree” algoritması uygulanır ve “Test & Score” modülüyle modelin başarımı ölçülür. Bu akış, verinin ham halinden sonuçların elde edilmesine kadar geçen tüm süreci temsil etmektedir. Görseldeki bağlantılar, veri ve model bileşenlerinin birbiriyle nasıl etkileşimde bulunduğunu açıkça göstermektedir.



**Şekil 3.2.** Orange Select Columns modülünde, veri setindeki nitelikler içinden hedef değişkenin (y) belirlenmesi ve kalan özneliklerin bağımsız değişken (features) olarak atanması.

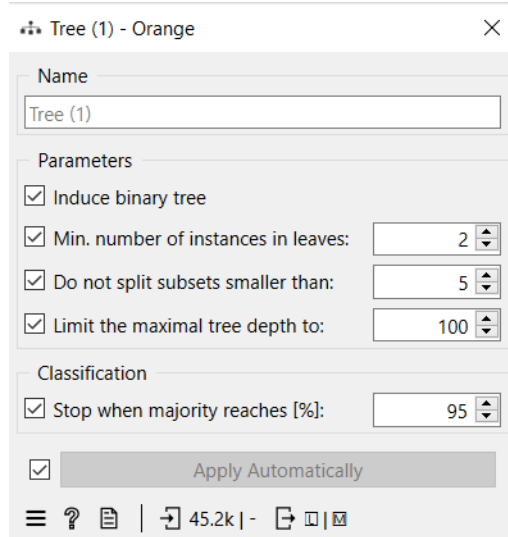
Bu şekil, verisetinde yer alan özneliklerin “bağımsız değişken” (features) ve “hedef değişken” (target) olarak nasıl ayrıldığını göstermektedir. Orange’da bu modül ile hangi değişkenlerin



modele girdi olarak verileceği (features) ve hangi değişkenin tahmin edilmeye çalışılacağı (target) belirlenir. Görselde, hedef değişken olarak y özniteliği seçilmiş; bu öznitelik müşterinin kampanyaya katılıp katılmadığını belirtir (yes/no). Diğer 16 öznitelik ise açıklayıcı değişkenler olarak sınıflandırılmıştır. Bu yapılandırma, modelin doğruluğunu doğrudan etkileyen kritik bir adımdır.

Orange Select Columns modülünde, veri setinde yer alan tüm değişkenler sınıflandırılarak bağımsız (features) ve bağımlı (target) değişkenler şeklinde tanımlanmıştır. Bu adım, modelin hangi değişkenleri giriş olarak alacağını ve hangi değişkeni tahmin etmeye çalışacağını belirlemek açısından kritik öneme sahiptir. Görselde, hedef değişken olan “y” alanı açıkça **Target** bölümüne aktarılmış; geri kalan 16 öznitelik ise **Features** olarak atanmıştır. Böylece model, müşterinin bankanın pazarlama kampanyasına katılım sağlayıp sağlamadığını (yes/no) tahmin edebilmek için yaş, eğitim, meslek, kredi bilgileri, iletişim tipi gibi çok sayıda değişkenden faydalanacak şekilde yapılandırılmıştır.

Bu modül sayesinde, sınıflandırma algoritması doğrudan hedef değişkene odaklanarak, giriş verisi üzerinden anlamlı örüntüler çıkarabilir. Ayrıca Orange arayüzünde yapılan bu net tanımlama, modelleme sürecinin şeffaf ve yeniden üretilebilir olmasını sağlar.



**Şekil 3.3.** Karar ağacı modeli için belirlenen parametre ayarları (yaprak sayısı, maksimum derinlik, bölünme sınırları)

Bu görselde, karar ağacı modelinin nasıl yapılandırıldığı görülmektedir. Parametreler, modelin nasıl dallanacağını belirler:

- **Min. number of instances in leaves:** Bir yaprakta yer alabilecek minimum örnek sayısı (2 olarak belirlenmiş).
- **Do not split subsets smaller than:** Bölünmeyecek minimum alt küme boyutu (5).
- **Limit tree depth:** Maksimum derinlik sınırı (100).

6

- **Stop when majority reaches %:** Bir sınıf belirli bir yüzdeye ulaştığında dal vermeyi durdurur (95%).  
Bu ayarlar, aşırı öğrenmeyi engellemeyi ve modelin genellenebilirliğini artırmayı

amaçlar. Bu yapılandırma, modeli dengeli ve açıklanabilir kılmak açısından son derece önemlidir.

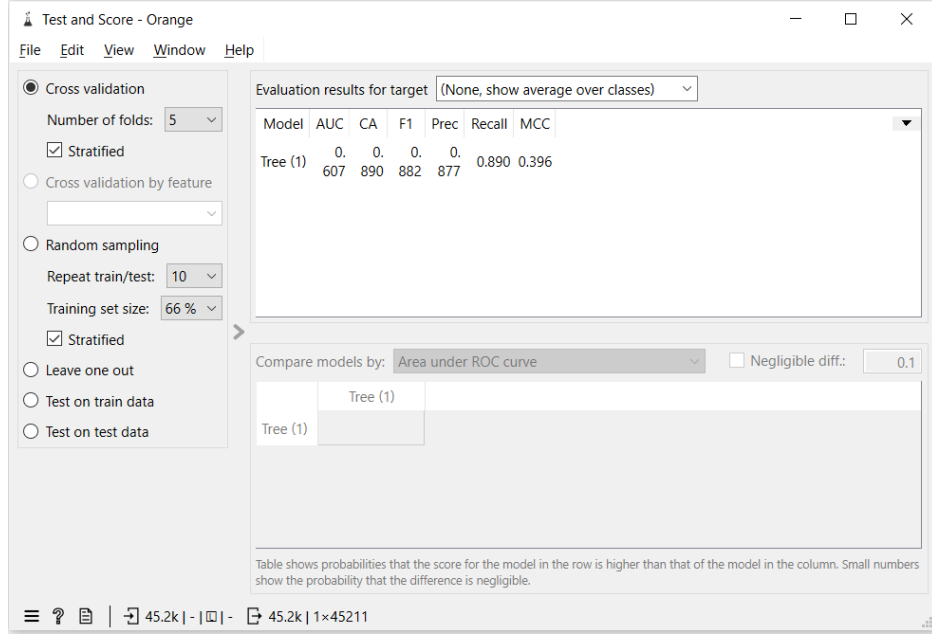
Bu projede, karar ağacı (Decision Tree) algoritması tercih edilerek UCI Bank Marketing veri seti üzerinde sınıflandırma modeli oluşturulmuştur. Karar ağacı algoritması, veri setindeki giriş değişkenlerini kullanarak hedef değişkenin (y: yes/no) öngörülmesini sağlar. Model, veriyi belirli karar kurallarına bölerek dallanma yapısı oluşturur ve her yaprakta bir karar noktasına ulaşılır. Bu yöntem hem yorumlanabilirlik hem de kolay uygulanabilirlik avantajı nedeniyle tercih edilmiştir.

Orange uygulamasında aşağıdaki akış ile model kurulmuştur:

- **CSV File Import:** Verinin içe aktarılması (noktalı virgül ayraç ile)
- **Select Columns:** Giriş değişkenlerinin (Features) ve hedef değişkenin (Target: y) belirlenmesi
- **Tree:** Karar ağacı modelinin kurulması
- **Tree Viewer:** Karar ağacının dallanma yapısını görselleştirme
- **Test & Score:** Modelin performans değerlendirmesinin yapılması

Karar ağacı yapısı incelendiğinde, ilk dallanmanın "**duration**" değişkeni üzerinden yapıldığı gözlemlenmiştir. Bu durum, kampanya süresi (görüşme uzunluğu) ile sonucun (evet/hayır) arasında doğrusal bir ilişki olduğunu göstermektedir. Ancak bu değişken, kampanyaya verilen cevap oluşuktan sonra oluştuğu için **veri kaçağı (data leakage)** olarak değerlendirilebilir. Bu nedenle bu tür değişkenlerin dikkatle ele alınması gerekmektedir.

#### 4. MODEL PERFORMANSI VE DEĞERLENDİRME



**Şekil 4.4.** Karar ağacı modeline ait test sonuçları ve metrik değerlendirme (CA, AUC, F1, Recall, Precision, MCC).

Bu şekil, oluşturulan karar ağacı modelinin başarımını ölçmek için kullanılan test sonuçlarını göstermektedir.

Orange platformunda **5-fold cross validation** yöntemi uygulanmıştır. Bu yöntem, verisini 5 parçaya bölerek her parça için eğitim ve test işlemi gerçekleştirir.

Orange platformunda 5 katlı çapraz doğrulama (5-fold cross validation) yöntemi ile değerlendirilen karar ağacı modelinin temel performans metrikleri aşağıdaki gibidir:

- **Accuracy (Doğruluk):** 0.890
- **F1 Skoru:** 0.882
- **Precision (Kesinlik):** 0.877
- **Recall (Duyarlılık):** 0.890
- **AUC (ROC Eğrisi Altı Alan):** 0.607
- **MCC (Matthews Korelasyon Katsayısı):** 0.396

Modelin genel doğruluk değeri oldukça yüksek (%89) çıkmıştır. Precision ve Recall değerlerinin birbirine yakın olması, modelin denge kurduğuna işaret etmektedir. Ancak AUC değerinin 0.607 olması, modelin sınıf ayrımında mükemmel bir performans göstermediğini ortaya koymaktadır. Bunun temel nedenlerinden biri, "duration" değişkeninin karar ağacının ilk dallanma noktasında yer alması ve bu özelliğin sonuca doğrudan etkide bulunması olabilir.

platformu üzerinden elde edilen doğruluk değeri de benzer şekilde %88 civarındadır. Bu durum, elde edilen sonuçların hem tutarlı hem de kullanılan yöntemin çalışma konusuna uygun olduğunu göstermektedir.

## 5. SONUÇ VE ÖNERİLER

Bu proje kapsamında, UCI Machine Learning Repository üzerinde yer alan Bank Marketing veri seti kullanılarak karar ağacı (Decision Tree) algoritması ile sınıflandırma modeli oluşturulmuş ve model Orange veri madenciliği platformu aracılığıyla uygulanmıştır. Veri setinde, bir bankanın doğrudan pazarlama kampanyasına yönelik müşteri bilgilerinin yer aldığı toplam 45.211 gözlem ve 17 değişken bulunmaktadır. Çalışmanın temel amacı, müşterilerin kampanyaya olumlu (yes) veya olumsuz (no) yanıt verip vermeyeceklerinin tahmin edilmesidir.

Karar ağacı algoritması tercih edilme sebebi; kolay anlaşılabilir yapısı, yorumlanabilirlik düzeyinin yüksek olması ve modelin şeffaf karar kuralları üretmesidir. Orange platformu, görsel iş akışı ile kullanıcıya modelleme sürecini sezgisel şekilde kurma ve analiz etme imkânı sunmaktadır. Bu bağlamda uygulanan modelleme sürecinde CSV dosyasının uygun şekilde içe aktarılması, bağımsız ve hedef değişkenlerin belirlenmesi, model parametrelerinin tanımlanması ve performans değerlendirme adımları dikkatle yürütülmüştür.

Model sonuçları incelendiğinde, karar ağacı algoritması %89 doğruluk (accuracy) oranına ulaşmıştır. Ayrıca F1-skora ait değer 0.882 olup modelin hem kesinlik (precision: 0.877) hem de duyarlılık (recall: 0.890) açısından dengeli bir sınıflandırma yaptığı görülmektedir. Bu sonuçlar, modelin kampanyaya olumlu dönüş yapacak bireyleri tanımlama açısından güçlü bir performans sergilediğini ortaya koymaktadır.

Ancak modelin AUC (Area Under Curve) değeri 0.607 olarak bulunmuştur. Bu değer, modelin sınıf ayrımı yapma becerisinin yüksek olmadığını, yani "evet" ve "hayır" sınıflarını ayırt etmede sınırlı bir performans sergilediğini göstermektedir. Ayrıca karar ağacı yapısı incelendiğinde, ilk ayrımın "duration" (görüşme süresi) değişkenine göre yapıldığı görülmektedir. Bu değişken, kampanyaya verilen cevaptan sonra ortaya çıkan bir ölçüt olduğu için, veri sızıntısı (data leakage) riski oluşturmaktadır. Modelin bu değişkene aşırı bağımlı olması, gerçek zamanlı tahminlerde modelin doğruluğunu düşürebilir. Bu nedenle, gelecekte yapılacak çalışmalarda "duration" değişkeninin modele dahil edilip edilmemesi dikkatle değerlendirilmelidir.

Bu bağlamda çalışmanın genel sonuçları şu şekilde özetlenebilir:

- Karar ağacı modeli, veri seti üzerinde yüksek doğruluk oranı ve tatmin edici F1-score değerleriyle başarılı sonuçlar üretmiştir.
- Modelde en belirleyici değişken "duration" olmuş; bu durum hem avantaj (doğruluk artırıcı) hem de sakınca (veri kaçağı riski) yaratmıştır.

- Orange platformu, karar ağacı gibi kural tabanlı modellerin uygulanması, görselleştirilmesi ve değerlendirilmesi açısından kullanıcı dostu ve işlevsel bir ortam sunmuştur.

Bu çalışmadan elde edilen bulgular ışığında, gelecek çalışmalara yönelik şu öneriler geliştirilebilir:

1. **Farklı algoritmalarla karşılaştırma yapılması:** Naive Bayes, Support Vector Machines (SVM), Neural Networks ve Random Forest gibi farklı sınıflandırma algoritmaları ile performans karşılaştırmaları yapılarak hangi yöntemin daha etkili sonuçlar verdiği test edilebilir.
2. **Özellik seçimi (Feature Selection) uygulamaları:** Tüm değişkenleri modele dahil etmek yerine, bilgi kazancı, Gini indeks veya Recursive Feature Elimination (RFE) gibi yöntemlerle en etkili öznitelikler seçilerek daha sade ve açıklayıcı modeller oluşturulabilir.
3. **Veri ön işleme çalışmalarının derinleştirilmesi:** "Unknown" değerleri içeren kategorik değişkenler yerine, eksik değerlerin çıkarılması, etiketlenmesi ya da impute edilmesiyle daha sağlıklı modeller üretilebilir.
4. **Data Leakage riskine karşı senaryo modellemeleri yapılması:** Özellikle “duration” gibi kampanya sonrası oluşan değişkenler dışlanarak model yeniden kurulabilir. Bu şekilde gerçek zamanlı tahminleme senaryoları test edilebilir.
5. **Zaman serisi analizi ve müşteri segmentasyonu yapılması:** Müşterilerin kampanyaya katılım zamanları veya geçmiş iletişim geçmişleri üzerinden zaman tabanlı analizler ve segmentasyon algoritmalarıyla derinlemesine içgörüler elde edilebilir.

Bu proje çalışması, veri madenciliği yöntemlerinin bankacılık ve pazarlama uygulamalarında nasıl etkili kullanılabileceğini göstermiştir. Doğru veri işleme adımları ve algoritma seçimiyle, müşteri davranışlarının tahmin edilmesi ve stratejik kararların verilmesi mümkün hale gelmektedir.

## 6. LİTERATÜR KARŞILAŞTIRMASI

Bu projede kullanılan veri seti, Elif Akkaya ve Safiye Turgay’ın “Unveiling the Power: A Comparative Analysis of Data Mining Tools through Decision Tree Classification on the Bank Marketing Dataset” başlıklı 2024 tarihli çalışmasında da kullanılmıştır. Söz konusu çalışmada, Orange başta olmak üzere KNIME, Weka ve RapidMiner gibi popüler veri madenciliği araçları karşılaştırmalı olarak analiz edilmiş; karar ağacı algoritması tüm platformlarda uygulanmıştır.

Akkaya ve Turgay’ın çalışmasında, Orange platformu ile uygulanan karar ağacı modelinden elde edilen doğruluk (accuracy) oranı %88,42 olarak rapor edilmiştir. Bu proje kapsamında geliştirilen modelde ise doğruluk oranı %89,0 olarak hesaplanmıştır. İki model arasında yaklaşık %0,6’lık bir fark bulunmakta olup, bu fark istatistiksel olarak küçük ancak dikkate değer sayılabilir.

Her iki çalışmanın benzer sonuçlar üretmesinin temel sebepleri arasında şunlar sayılabilir:

- Aynı veri setinin kullanılması
- Aynı algoritmanın uygulanması (karar ağacı)
- Orange platformunun tercih edilmesi (benzer parametreler ve çapraz doğrulama)

<b>Karşılaştırma Unsuru</b>	<b>Akkaya &amp; Turgay (2024)</b>	<b>Simge Söyler)</b>
Doğruluk Oranı	%88,42	%89,00
Platform	Orange	Orange
Veri Ön İşleme	Detay belirtilmemiş	Select Columns ile duration dahil, y net hedef atanmış
Değişken Kullanımı	Muhtemelen tüm değişkenler kullanıldı	Duration dahil, veri kaçağı riski analiz edildi
Yorumlanabilirlik	Yüzeysel model analizi	Tree Viewer ile değişken etkisi derinlemesine incelendi

Bu projenin avantajlı yönlerinden biri, “duration” değişkeninin model üzerindeki etkisinin veri kaçağı (data leakage) açısından eleştirel biçimde tartışılmasıdır. Akkaya ve Turgay’ın çalışmasında bu noktaya değinilmemiştir. Ayrıca, bu projede yapılan detaylı Select Columns yapılandırması ve model metriklerinin tek tek değerlendirilmesi, uygulama kalitesini artırmıştır.

Sonuç olarak, farklı araştırmalarda benzer modellerin benzer performanslar göstermesi, veri setinin karakteristik yapısına, algoritmanın doğasına ve araçların benzer işleyiş mantığına dayanmaktadır. Ancak bu proje, yorumlanabilirlik, görselleştirme ve metodolojik şeffaflık açısından daha uygulamaya yönelik derinlik sunmaktadır.

## 7. KAYNAKÇA

UCI Machine Learning Repository. (n.d.). *Bank marketing dataset*. University of California, Irvine. Retrieved May 25, 2025, from <https://archive.ics.uci.edu/dataset/222/bank+marketing>