



## **TÜBİTAK–2209-A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI**

**Başvuru formunun Arial 9 yazı tipinde, her bir konu başlığı altında verilen açıklamalar göz önünde bulundurularak hazırlanması ve ekler hariç toplam 20 sayfayı geçmemesi beklenir (Alt sınır bulunmamaktadır). Değerlendirme araştırma önerisinin özgün değeri, yöntemi, yönetimi ve yaygın etkisi başlıkları üzerinden yapılacaktır.**

### **ARAŞTIRMA ÖNERİSİ FORMU**

2024 Yılı

... Dönem Başvurusu

**2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI**  
**ARAŞTIRMA ÖNERİSİ FORMU**

**A. GENEL BİLGİLER**

<b>Başvuru Sahibinin Adı Soyadı: Ebubekir Sıddık Nazlı</b>
<b>Araştırma Önerisinin Başlığı: Veri Madenciliği ile Dolandırıcılık Tespiti</b>
<b>Danışmanın Adı Soyadı: Selman Yakut</b>
<b>Araştırmanın Yürütüleceği Kurum/Kuruluş: İnönü Üniversitesi Mühendislik Fakültesi</b>

**ÖZET**

Türkçe özetin araştırma önerisinin (a) özgün değeri, (b) yöntemi, (c) yönetimi ve (d) yaygın etkisi hakkında bilgileri kapsamı beklenir. Bu bölümün en son yazılması önerilir.

**Özet**

Bu çalışma, veri madenciliği yöntemleriyle dolandırıcılık tespitinde yenilikçi yaklaşımlar geliştirmeyi amaçlamaktadır. Araştırmanın özgün değeri, finans sektöründe büyük kayıplara neden olan dolandırıcılık faaliyetlerini, güçlü makine öğrenimi algoritmaları ve veri madenciliği teknikleri kullanarak daha etkili bir şekilde tespit edebilme potansiyeline dayanır. Çalışmada, rastgele ormanlar, destek vektör makineleri (SVM), lojistik regresyon, XGBoost ve k-en yakın komşu (KNN) gibi önde gelen algoritmalar, gerçek ve simüle edilmiş veri kümeleri üzerinde uygulanmıştır.

Yöntem olarak, dengesiz veri kümeleri için SMOTE gibi veri artırma teknikleri kullanılmış, sınıflandırma performansını artırmak için çapraz doğrulama ve ROC-AUC gibi değerlendirme metriklerine yer verilmiştir. Araştırmanın yönetimi, algoritmaların etkinliğini kıyaslayan sistematik bir model geliştirme ve test sürecini kapsamıştır. Çalışmada kullanılan modeller, sahte işlemleri yüksek doğruluk oranıyla tespit etmeye odaklanmıştır.

Yaygın etkisi bakımından, bu çalışma, dolandırıcılığın erken tespiti için kullanılabilecek etkili bir çözüm sunmakta ve finansal kurumların güvenlik süreçlerini iyileştirmelerine katkı sağlamaktadır. Bu yöntemlerin, sadece finans sektöründe değil, diğer sahteciliğe açık alanlarda da uygulanabilirliği bulunmaktadır.

**Anahtar Kelimeler:** Veri madenciliği, dolandırıcılık tespiti, makine öğrenimi, sınıflandırma algoritmaları, finansal güvenlik.

**1. ÖZGÜN DEĞER**

**1.1. Konunun Önemi, Araştırma Önerisinin Özgün Değeri ve Araştırma Sorusu/Hipotezi**

Araştırma önerisinde ele alınan konunun kapsamı ve sınırları ile önemi literatürün eleştirel bir değerlendirmesinin yanı sıra nitel veya nicel verilerle açıklanır.

Özgün değer yazılırken araştırma önerisinin bilimsel değeri, farklılığı ve yeniliği, hangi eksikliği nasıl gidereceği veya hangi soruna nasıl bir çözüm geliştireceği ve/veya ilgili bilim veya teknoloji alan(lar)ına kavramsal, kuramsal ve/veya metodolojik olarak ne gibi özgün katkılarda bulunacağı literatüre atıf yapılarak açıklanır.

Önerilen çalışmanın araştırma sorusu ve varsa hipotezi veya ele aldığı problem(ler)i açık bir şekilde ortaya konulur.

Dolandırıcılık, özellikle finans sektörü gibi yüksek risk içeren alanlarda büyük ekonomik kayıplara yol açmaktadır. Literatürde, dolandırıcılık tespitine yönelik birçok algoritma ve yöntem önerilmesine rağmen, dengesiz veri kümeleri, concept drift (zamanla değişen veri yapıları) ve dolandırıcılık işlemlerinin etiketlenmesindeki gecikmeler gibi problemler hala önemli bir zorluk olarak kalmaktadır. Bu nedenle, dolandırıcılık tespitinde daha etkili ve genelleştirilebilir modellerin geliştirilmesine ihtiyaç duyulmaktadır. Araştırmanın özgün değeri, mevcut yöntemlerin eksikliklerini gidermeye yönelik yeni bir yaklaşım sunmasında yatmaktadır. Bu çalışmada, rastgele ormanlar, destek vektör makineleri (SVM), lojistik regresyon, XGBoost ve k-en yakın komşu (KNN) algoritmalarının performansları, dengesiz veri kümelerine özel teknikler ve çapraz doğrulama yöntemleri ile birleştirilerek karşılaştırılacaktır. Ayrıca, SMOTE tekniği kullanılarak küçük sınıfların dengelemesi sağlanacak ve böylece sınıflandırma performansı artırılacaktır. Çalışma, dolandırıcılık tespitinde

## 2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI ARAŞTIRMA ÖNERİSİ FORMU

yüksek doğruluk ve duyarlılık oranlarına ulaşmayı hedefleyerek literatüre metodolojik bir katkı sağlayacaktır.

Bu araştırmanın temel sorusu şudur: **"Dengesiz veri kümelerinde dolandırıcılık işlemlerini tespit etmek için hangi makine öğrenimi algoritması en yüksek performansı göstermektedir?"**. Hipotez ise şu şekilde formüle edilmiştir: **"SMOTE ile dengelenmiş veri kümelerinde XGBoost algoritması, diğer algoritmalara kıyasla daha yüksek doğruluk ve F1 skoruna ulaşacaktır."**

Araştırma, mevcut algoritmaların sınırlılıklarını ele alarak, özellikle finansal dolandırıcılık gibi kritik alanlarda daha etkili ve genelleştirilebilir çözümler geliştirilmesine olanak tanıyacaktır. Ayrıca, önerilen yöntemlerin diğer sektörlerdeki sahtecilik tespitine de uyarlanabilirliği, bu çalışmanın hem bilimsel hem de pratik katkısını artıracaktır.

### 1.2. Amaç ve Hedefler

Araştırma önerisinin amacı ve hedefleri açık, ölçülebilir, gerçekçi ve araştırma süresince ulaşılabilir nitelikte olacak şekilde yazılır.

Bu araştırmanın amacı, dolandırıcılık tespiti için makine öğrenimi ve veri madenciliği tekniklerini kullanarak yüksek doğruluk oranına sahip, genelleştirilebilir bir model geliştirmektir. Araştırma, dengesiz veri kümelerinde dolandırıcılık işlemlerini etkili bir şekilde sınıflandırmak ve bu süreçte SMOTE gibi veri artırma yöntemlerinin, rastgele ormanlar, destek vektör makineleri (SVM), lojistik regresyon, XGBoost ve k-en yakın komşu (KNN) algoritmaları üzerindeki etkisini incelemeyi hedeflemektedir.

Hedefler şu şekilde belirlenmiştir:

1. Dengesiz veri kümeleri için SMOTE gibi veri işleme yöntemlerini uygulayarak sınıflandırma algoritmalarının performansını optimize etmek.
2. Rastgele ormanlar, SVM, lojistik regresyon, XGBoost ve KNN algoritmalarını sahte ve gerçek işlemleri tespit etmede karşılaştırmak.
3. Doğruluk, kesinlik, duyarlılık, F1 skoru ve ROC-AUC gibi performans metriklerini analiz ederek en iyi sonuç veren algoritmayı belirlemek.
4. Önerilen modelin finans sektöründeki dolandırıcılık tespitine katkısını ortaya koymak ve diğer sektörlerde uygulanabilirliğini değerlendirmek.

Araştırmanın çıktısı, dolandırıcılıkla mücadele süreçlerinde uygulanabilecek etkili bir model sunarken, makine öğrenimi algoritmalarının dengesiz veri kümelerindeki performansları hakkında da kapsamlı bir analiz sağlayacaktır. Bu hedefler, hem pratikte hem de akademik alanda kullanılabilir, somut bir çözüm geliştirilmesine yönelik açık ve ölçülebilir bir çerçeve sunmaktadır.

### 2. YÖNTEM

Araştırma önerisinde uygulanacak yöntem ve araştırma teknikleri (veri toplama araçları ve analiz yöntemleri dahil) ilgili literatüre atıf yapılarak açıklanır. Yöntem ve tekniklerin çalışmada öngörülen amaç ve hedeflere ulaşmaya elverişli olduğu ortaya konulur.

Yöntem bölümünün araştırmanın tasarımı, bağımlı ve bağımsız değişkenleri ve istatistiksel yöntemleri kapsamı gerekir. Araştırma önerisinde herhangi bir ön çalışma veya fizibilite yapıldıysa bunların sunulması beklenir. Araştırma önerisinde sunulan yöntemlerin iş paketleri ile ilişkilendirilmesi gerekir.

Bu araştırma, dolandırıcılık tespiti için veri madenciliği ve makine öğrenimi tekniklerini kullanarak sistematik bir analiz gerçekleştirmeyi amaçlamaktadır. Çalışmada kullanılacak yöntemler ve teknikler, literatürde kanıtlanmış yaklaşımlara dayanmaktadır ve bu araştırmanın amaç ve hedeflerine ulaşmada uygun bir çerçeve sunmaktadır.

#### Araştırma

#### Tasarımı:

Araştırma, deneysel bir tasarıma sahiptir ve dengesiz veri kümeleri üzerinde makine öğrenimi algoritmalarının performansını karşılaştırmaya odaklanır. Bağımlı değişken, dolandırıcılık tespiti başarı oranıdır (doğruluk, kesinlik, duyarlılık gibi metrikler). Bağımsız değişkenler ise kullanılan algoritmalar (rastgele ormanlar, SVM, lojistik regresyon, XGBoost ve KNN) ve veri işleme teknikleridir (SMOTE gibi).

#### Veri

#### Toplama

#### Araçları:

Araştırma, kamuya açık veri setlerinden Kaggle'dan elde edilen finansal işlem verilerini kullanacaktır. Bu veri setleri, sahte ve gerçek işlemleri içeren etiketlenmiş veri noktalarından oluşmaktadır. Veri toplama sürecinde, dengesizlik problemi göz önünde bulundurularak, sınıf dağılımları analiz edilecektir.

#### Analiz Yöntemleri:

1. **Veri Ön İşleme:** Veri seti, eksik değerlerin giderilmesi, anormal değerlerin kontrol edilmesi ve kategorik değişkenlerin sayısal forma dönüştürülmesi gibi adımlardan geçirilecektir.

**2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI**  
**ARAŞTIRMA ÖNERİSİ FORMU**

2. **Veri Dengesi:** Dengesiz veri kümelerinde, SMOTE (Synthetic Minority Over-sampling Technique) yöntemi kullanılarak küçük sınıflar artırılacaktır.
3. **Model Eğitimi:** Rastgele ormanlar, SVM, lojistik regresyon, XGBoost ve KNN algoritmaları, eğitim veri seti üzerinde çapraz doğrulama kullanılarak eğitilecektir.
4. **Performans Değerlendirme:** Modeller, test veri seti üzerinde doğruluk, kesinlik, duyarlılık, F1 skoru ve ROC-AUC metrikleriyle karşılaştırılacaktır.

**İstatistiksel**

**Yöntemler:**

Analiz sürecinde çapraz doğrulama ve hipotez testleri uygulanarak, algoritmaların performans farklılıklarının istatistiksel olarak anlamlı olup olmadığı değerlendirilecektir. Ayrıca, karar ağaçları ve önem dereceleri kullanılarak algoritmaların hangi özellikleri daha önemli gördüğü belirlenecektir.

**İş Paketleri ve Öngörülen Aşamalar:**

1. Veri setinin toplanması ve ön işlenmesi.
2. Veri dengesizlik sorunlarının çözülmesi.
3. Makine öğrenimi algoritmalarının eğitimi ve testi.
4. Performans değerlendirme ve sonuçların raporlanması.

Bu yöntemler, araştırmanın hedeflerine ulaşılmasını sağlayacak ve literatürdeki mevcut çalışmalara hem teorik hem de uygulamalı katkılar sunacaktır.

## 3 PROJE YÖNETİMİ

## 3.1 İş- Zaman Çizelgesi

Araştırma önerisinde yer alacak başlıca iş paketleri ve hedefleri, her bir iş paketinin hangi sürede gerçekleştirileceği, başarı ölçütü ve araştırmanın başarısına katkısı “İş-Zaman Çizelgesi” doldurularak verilir. Literatür taraması, gelişme ve sonuç raporu hazırlama aşamaları, araştırma sonuçlarının paylaşımı, makale yazımı ve malzeme alımı ayrı birer iş paketi olarak gösterilmemelidir.

Başarı ölçütü olarak her bir iş paketinin hangi kriterleri sağladığında başarılı sayılacağı açıklanır. Başarı ölçütü, ölçülebilir ve izlenebilir nitelikte olacak şekilde nicel veya nitel ölçütlerle (ifade, sayı, yüzde, vb.) belirtilir.

## İŞ-ZAMAN ÇİZELGESİ (\*)

İP No	İş Paketlerinin Adı ve Hedefleri	Kim(ler) Tarafından Gerçekleştirileceği	Zaman Aralığı (09-06 Ay)	Başarı Ölçütü ve Projenin Başarısına Katkısı
1	İhtiyaç Analizi ve Planlama	Ebubekir Sıddık Nazlı	09-09	Gereksinim dokümanının tamamlanması. Doğru analizle projeye sağlam bir temel oluşturulması.
2	Tasarım ve Mimari Oluşturma	Ebubekir Sıddık Nazlı	09-10	Tasarım dokümanının onaylanması. Proje mimarisinin kullanıcı ihtiyaçlarını karşılayacak şekilde hazırlanması.
3	Geliştirme ve Kodlama	Ebubekir Sıddık Nazlı	10-12	Modüllerin başarılı bir şekilde geliştirilmesi ve test edilmesi. İşlevselliklerin eksiksiz şekilde hayata geçirilmesi.
4	Test ve Hata Giderme	Ebubekir Sıddık Nazlı	12-01	Tüm testlerin başarılı bir şekilde tamamlanması ve hataların giderilmesi. Sorunsuz bir teslimat yapılması.
5	Kullanıcı Eğitim ve Destek Hizmetleri	Ebubekir Sıddık Nazlı	01-02	Eğitimlerin tamamlanması ve kullanıcıların sistemi etkin bir şekilde kullanabilmesi. Projenin kullanıcı dostu olması sağlanır.
6	Canlıya Alma ve Takip	Ebubekir Sıddık Nazlı	02-06	Sistemin başarıyla devreye alınması ve performans takibinin yapılması. Projenin sürdürülebilirliği ve verimliliği artırılır.

(\*) Çizelgedeki satırlar ve sütunlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

**2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI**  
**ARAŞTIRMA ÖNERİSİ FORMU**

### 3.2 Risk Yönetimi

Araştırmanın başarısını olumsuz yönde etkileyebilecek riskler ve bu risklerle karşılaşıldığında araştırmanın başarıyla yürütülmesini sağlamak için alınacak tedbirler (B Planı) ilgili iş paketleri belirtilerek ana hatlarıyla aşağıdaki Risk Yönetimi Tablosu'nda ifade edilir. B planlarının uygulanması araştırmanın temel hedeflerinden sapmaya yol açmamalıdır.

**RİSK YÖNETİMİ TABLOSU\***

İP No	En Önemli Riskler	Risk Yönetimi (B Planı)
1	Veri setinin yetersiz kalması veya veri dengesizliği sorunlarının çözülmemesi	Kamuya açık farklı veri setleri kullanılarak veri çeşitliliği artırılacaktır. Veri artırma yöntemleri (örneğin SMOTE veya GAN) uygulanacaktır.
2	Algoritmaların performansının beklenen düzeye ulaşamaması	Parametre optimizasyonu yapılacak, hiperparametre arama teknikleri (Grid Search veya Random Search) uygulanacaktır. Alternatif algoritmalar (örneğin, LightGBM veya CatBoost) test edilecektir.
3	Veri ön işleme aşamasında karşılaşılan eksik veya hatalı veriler	Eksik veri doldurma teknikleri (örneğin, ortalama veya medyan ile tamamlama) ve anormal değerlerin çıkarılması gibi yöntemler uygulanacaktır.
4	Çalışma sürecinde yazılım veya donanım kaynaklı teknik aksaklıklar	Çalışmalar düzenli olarak yedeklenecek ve bulut tabanlı platformlar (Google Colab, AWS veya Azure) yedek seçenek olarak kullanılacaktır.
5	Performans değerlendirme aşamasında istatistiksel sonuçların anlamlılık göstermemesi	Veri seti boyutu artırılarak analizler tekrar yapılacak, farklı değerlendirme metrikleri (Precision-Recall eğrisi gibi) kullanılacaktır.

(\*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

### 3.3. Araştırma Olanakları

Bu bölümde projenin yürütüleceği kurum ve kuruluşlarda var olan ve projede kullanılacak olan altyapı/ekipman (laboratuvar, araç, makine-teçhizat, vb.) olanakları belirtilir.

**ARAŞTIRMA OLANAKLARI TABLOSU (\*)**

Kuruluştaki Bulunan Altyapı/Ekipman Türü, Modeli (Laboratuvar, Araç, Makine-Teçhizat, vb.)	Projede Kullanım Amacı
Yazılım Lisansları (Pardus, Microsoft)	İşletim Sistemi ve Microsoft araçları
Yüksek Hızlı İnternet Ağı	Veri aktarımı ve uzaktan erişim sağlanması
Kütüphane ve Bilgisayar laboratuvarları	Çalışma Alanı

(\*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

### 4. YAYGIN ETKİ

Önerilen çalışma başarıyla gerçekleştirildiği takdirde araştırmadan elde edilmesi öngörülen ve beklenen yaygın etkilerin neler olabileceği, diğer bir ifadeyle yapılan araştırmadan ne gibi çıktı, sonuç ve etkilerin elde edileceği aşağıdaki tabloda verilir.

**ARAŞTIRMA ÖNERİSİNDEN BEKLENEN YAYGIN ETKİ TABLOSU**

Yaygın Etki Türleri	Önerilen Araştırmadan Beklenen Çıktı, Sonuç ve Etkiler
---------------------	--

**2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI**  
**ARAŞTIRMA ÖNERİSİ FORMU**

<b>Bilimsel/Akademik</b> (Makale, Bildiri, Kitap Bölümü, Kitap)	- Dergilerde yayımlanacak 1 adet makale. - Alanla ilgili bir kitap bölümünün oluşturulması.
<b>Ekonomik/Ticari/Sosyal</b> (Ürün, Prototip, Patent, Faydalı Model, Üretim İzni, Çeşit Tescili, Spin-off/Start-up Şirket, Görsel/İşitsel Arşiv, Envanter/Veri Tabanı/Belgeleme Üretimi, Telif Konu Olan Eser, Medyada Yer Alma, Fuar, Proje Pazarı, Çalıştay, Eğitim vb. Bilimsel Etkinlik, Proje Sonuçlarını Kullanacak Kurum/Kuruluş, vb. diğer yaygın etkiler)	- İlgili sektörlerde kullanılabilecek yeni bir algoritma yöntemi - Yerel medya ve bilimsel etkinliklerde proje tanıtımı.
<b>Araştırmacı Yetiştirilmesi ve Yeni Proje(ler) Oluşturma</b> (Yüksek Lisans/Doktora Tezi, Ulusal/Uluslararası Yeni Proje)	- 1 adet yüksek lisans tezine katkı sağlanması. - Araştırma yazısına yardımcı kaynak olması.

**5. BÜTÇE TALEP ÇİZELGESİ**

Bütçe Türü	Talep Edilen Bütçe Miktarı (TL)	Talep Gerekçesi
<b>Sarf Malzeme</b>	86.000 TL	Deneysel çalışmalar için gerekli malzemeler (Laptop ve Veri Madenciliği için ekran kartı).
<b>Makina/Teçhizat (Demirbaş)</b>	50.000 TL	Özel analiz cihazlarının ve test ekipmanlarının temini.
<b>Hizmet Alımı</b>	240.000 TL	Kendi çalışmamın hakkı olarak gördüğüm ücret
<b>Ulaşım</b>	17.000 TL	Gerekli merkezlere gezi
<b>TOPLAM</b>	393.000 TL	Tüm Masraflar

**NOT:** Bütçe talebiniz olması halinde hem bu tablonun hem de TÜBİTAK Yönetim Bilgi Sistemi (TYBS) başvuru ekranında karşınıza gelecek olan bütçe alanlarının doldurulması gerekmektedir. Yukardaki tabloda girilen bütçe kalemlerindeki rakamlar ile, TYBS başvuru ekranındaki rakamlar arasında farklılık olması halinde TYBS ekranındaki veriler dikkate alınır ve başvuru sonrasında değiştirilemez.

**6. BELİRTMEK İSTEDİĞİNİZ DİĞER KONULAR**

Sadece araştırma önerisinin değerlendirilmesine katkı sağlayabilecek bilgi/veri (grafik, tablo, vb.) eklenebilir.

Yok.

**7. EKLER**

**EK-1: KAYNAKLAR**