

**AYGAZ**  
**YAPAY ZEKAYA GİRİŞ**  
**PROJE SUNUMU**

**YUSUF KILIÇ**

# 1. GİRİŞ

---

Bu rapor, MNIST veri seti üzerinde yapılan makine öğrenimi projesini detaylandırmaktadır. Projenin amacı, farklı sınıflandırma modellerini kullanarak el yazısı rakamlarını tanıma görevinde en etkili modeli belirlemektir. Elde edilen sonuçlar, farklı makine öğrenimi tekniklerinin karşılaştırılması ve performanslarının değerlendirilmesi yoluyla sunulmaktadır.



## 2. VERİ SETİ HAKKINDA BİLGİLER VE VERİ SETİ HAZIRLIĞI

---

MNIST veri seti, Modified National Institute of Standards and Technology (MNIST) tarafından hazırlanan ve 0 ile 9 arasındaki rakamları içeren bir veri setidir. Veri seti, 60,000 eğitim ve 10,000 test görüntüsünden oluşmaktadır. Her bir görüntü, 28x28 piksel boyutundadır ve siyah-beyaz (grayscale) formatında kodlanmıştır.

Proje kapsamında, MNIST veri seti öncelikle küçültülmüş ve işlenmiş bir alt kümesi kullanılmıştır. Eğitim setinden 5,000 görüntü ve test setinden 1,000 görüntü seçilmiştir. Seçilen görüntüler normalize edilmiş ve 0 ile 1 arasında ölçeklenmiştir.

### 3. KULLANILAN MODELLER

---

Proje, altı farklı sınıflandırma modeli üzerinde çalışmıştır:

- **K-Nearest Neighbors (KNN):** Komşuluk temelli sınıflandırma yöntemi.
- **Support Vector Machine (SVM):** Optimal hiperdüzlemi bulmaya çalışan bir sınıflandırma yöntemi.
- **Logistic Regression:** İkili ve çok sınıflı sınıflandırma problemleri için kullanılan lineer bir model.
- **Decision Tree:** Karar ağacı yapısını kullanarak sınıflandırma yapan bir model.
- **Random Forest:** Birden fazla karar ağacının bir araya gelmesiyle oluşturulan bir model.
- **Gradient Boosting:** Zayıf öğrenicilerin bir araya gelmesiyle güçlü bir model oluşturan bir ensemble yöntemi.

## 4. KULLANILAN TEKNİKLER

---

### Veri Ön İşleme

- Veri ön işleme adımları, veri setinin hazırlanması ve modele veri sağlanmadan önce yapılan işlemleri kapsar. Bu projede MNIST veri seti üzerinde şu ön işleme adımları uygulanmıştır:
- **Veri Küçültme:** Eğitim setinden 5,000 görüntü ve test setinden 1,000 görüntü seçilmiştir. Bu işlem, model eğitimi hızlandırmak ve denemeleri yönetilebilir kılmak amacıyla yapılmıştır.
- **Görüntü Normalizasyonu:** Görüntüler, 0 ile 255 arasında değişen piksel değerlerinden normalize edilerek 0 ile 1 arasında ölçeklendirilmiştir. Bu adım, modelin eğitimi stabilize etmek ve daha hızlı bir şekilde yakınsama sağlamak için yapılmıştır.
- **Görüntü Düzleştirme (Flatten):** Görüntüler, 28x28 piksel boyutundan 784 piksel uzunluğunda vektörlere dönüştürülmüştür. Bu işlem, her bir pikselin ayrı bir özellik olarak ele alınmasını sağlayarak sınıflandırma modellerinin kullanımını kolaylaştırmıştır.



## 5. DEĞERLENDİRME METRİKLERİ

---

Her model için eğitim ve değerlendirme süreçlerinde aşağıdaki değerlendirme metrikleri kullanılmıştır:

- **Doğruluk (Accuracy):** Doğru tahmin edilen örneklerin toplam örnekler içindeki oranını ifade eder.
- **Kesinlik (Precision):** Pozitif olarak tahmin edilen örneklerin gerçekten pozitif olan örneklerin oranını ifade eder. Çok sınıflı sınıflandırmada weighted average olarak hesaplanabilir.
- **ROC AUC:** Receiver Operating Characteristic (ROC) eğrisi altında kalan alanın (AUC) çok sınıflı sınıflandırmalar için genellikle "one-vs-one" veya "one-vs-rest" stratejileriyle hesaplanan bir metriktir. Sınıflandırıcıların sınıflandırma performansını değerlendirmek için kullanılır.

---

Bu yöntemler, projenin başlangıcından sonuna kadar kullanılmış ve sonuçların analiz edilmesi için temel oluşturmuştur. Her bir adım, veri setinin hazırlanmasından model eğitime ve sonuçların değerlendirilmesine kadar sistematik bir şekilde uygulanmıştır. Bu sayede elde edilen bulgular, farklı sınıflandırma tekniklerinin performansını objektif bir şekilde karşılaştırmamıza olanak tanımıştır.

## 6. TEMEL BULGULAR

---

Proje sonucunda elde edilen temel bulgular şunlardır:

- SVM ve Lojistik Regresyon modelleri, diğer modellere kıyasla daha yüksek doğruluk skorları elde etmiştir.
- Karar Ağacı ve Rastgele Orman modelleri ise daha düşük performans göstermiştir.
- Gradient Boosting modeli ise orta düzeyde bir performans sergilemiştir.
- En iyi model olarak SVM seçilmiş ve test seti üzerinde de başarılı sonuçlar vermiştir.