# **Churn Proje Yol Haritasi**

### **1. Keşifsel Veri Analizi (EDA)**

* **Eksik Veri Kontrolü:** Eksik değerlerin olup olmadığını kontrol edin ve bu eksik değerleri nasıl işleyeceğinize karar verin.
* **Veri Dağılımları:** Her sütunun (örneğin, yaş, kredi puanı, maaş) dağılımını görselleştirin.
* **Korelasyon Analizi:** Sayısal değişkenlerin birbirleriyle olan korelasyonlarını inceleyin.
* **Gruplama:** Churn eden ve etmeyen müşteriler arasındaki farkları (örneğin, yaş, kredi puanı, gelir) analiz edin.
* **Kategorik Değişkenlerin Analizi:** Cinsiyet, coğrafya gibi kategorik değişkenlerin müşteri kaybına etkisini analiz edin.

### **2. Veri Ön İşleme**

* **Kategorik Değişkenlerin Kodlanması:** Geography, Gender, HasCrCard gibi kategorik değişkenleri sayısal değerlere çevirin.
* **Ölçekleme:** Kredi puanı, maaş gibi sayısal değişkenleri ölçekleyin (özellikle makine öğrenimi modellerinde kullanılacaksa).
* **Özellik Seçimi:** Analizde kullanılacak en etkili değişkenleri belirleyin.

### **3. Görselleştirme**

* **Churn Oranı Görselleştirme:** Kaç müşterinin kaybedildiğini (Exited sütunu) pasta grafiği veya çubuk grafiğiyle gösterin.
* **Coğrafyaya Göre Churn Oranı:** Hangi ülkelerde müşteri kaybının daha yüksek olduğunu inceleyin.
* **Yaş ve Kredi Puanı Dağılımı:** Scatter plot ile müşteri kaybı (Exited) ile ilişkisini görselleştirin.

### **4. Makine Öğrenimi Modelleri**

* **Hedef Değişken:** Exited sütunu, müşterinin kaybedip kaybetmediğini gösteren bir sınıflandırma problemi sunuyor.
* **Model Seçimi:**
  + **Lojistik Regresyon**
  + **Karar Ağaçları ve Random Forest**
  + **XGBoost veya LightGBM**
  + **Destek Vektör Makineleri (SVM)**
* **Model Değerlendirme:**
  + Confusion Matrix, Accuracy, Precision, Recall, F1 Score gibi metriklerle değerlendirme yapın.
  + ROC AUC eğrisi ile performansı inceleyin.

### **5. Model Optimizasyonu**

* **Hyperparameter Tuning:** GridSearchCV veya RandomSearchCV kullanarak en iyi parametreleri bulun.
* **Feature Engineering:** Yeni değişkenler türetin (örneğin, Kredi puanı ve yaş oranı).

### **6. Sonuçların Sunumu ve Aksiyon Önerileri**

* Hangi müşteri gruplarının daha fazla kaybedildiğini belirleyin ve bu gruplar için öneriler sunun.
* Sadık müşterileri elde tutmak için stratejiler geliştirin.

### **Makine Öğrenmesi Adımları:**

### **1. Hedef Belirleme ve Problem Tanımı**

* **Hedef:** Exited sütununu tahmin etmek (Bu, müşterinin bankadan ayrılıp ayrılmadığını gösteren **ikili sınıflandırma** problemidir).
* **Girdi Özellikleri:** Yaş, kredi puanı, gelir gibi değişkenleri kullanarak churn olup olmayacağını tahmin edeceğiz.

### **2. Veri Hazırlığı ve Ön İşleme**

* **Eksik Verilerin Kontrolü:** Eksik değerler varsa bunları uygun şekilde doldur veya temizle.
* **Kategorik Değişkenlerin Kodlanması:**
  + Geography, Gender, HasCrCard, IsActiveMember gibi kategorik değişkenleri **Label Encoding** veya **One-Hot Encoding** ile sayısal hale getirin.
* **Veri Ölçekleme:** Özellikle lojistik regresyon veya destek vektör makineleri gibi algoritmalar için sayısal verileri ölçeklemek iyi bir fikirdir.
  + Standartlaştırma: Z-score yöntemi veya Min-Max ölçekleme kullanılabilir.
* **Veri Bölme:**
  + Veriyi eğitim ve test setlerine bölün (örn: %80 eğitim - %20 test).

### **3. Makine Öğrenmesi Algoritmalarının Seçimi**

Kullanabileceğiniz bazı algoritmalar şunlardır:

* **Lojistik Regresyon:** Basit ve yorumlanabilir bir model.
* **Karar Ağaçları:** Kuralları anlamak için kullanışlıdır.
* **Random Forest:** Birden fazla karar ağacıyla daha iyi performans sağlar.
* **XGBoost veya LightGBM:** Büyük veri setlerinde çok güçlüdür.
* **Destek Vektör Makineleri (SVM):** Özellikle küçük ve karmaşık veri setlerinde etkilidir.

### **4. Model Eğitme ve Değerlendirme**

* **Performans Metrikleri:**
  + Accuracy (Doğruluk)
  + Precision (Kesinlik)
  + Recall (Duyarlılık)
  + F1 Skoru
  + ROC AUC Skoru

### **5. Model Optimizasyonu**

* **Hyperparameter Tuning:** GridSearchCV veya RandomSearchCV ile en iyi parametreleri arayın.
* **Feature Selection:** En önemli özellikleri belirleyerek gereksiz olanları çıkarabilirsiniz.

### **6. Model Test Etme ve Tahmin Yapma**

* Test verisiyle model performansını ölçün.
* Gerçek ve tahmin edilen sonuçları karşılaştırarak modelin başarısını analiz edin.

### **7. Sonuçların Sunumu**

* Modele göre hangi özelliklerin müşteri kaybını daha çok etkilediğini açıklayın.
* Aksiyon önerileri geliştirin.