

Müşteri Kaybı (Churn) Tahmin Projesi Dokümanı

Projenin Amacı

Bu projenin temel amacı, bir telekomünikasyon şirketinin mevcut müşterilerinden hangilerinin hizmeti bırakma (churn) eğiliminde olduğunu önceden tahmin etmektir. Böylece şirket, kaybetme riski taşıyan müşterilere özel stratejiler geliştirerek sadakatı artırabilir.

Veri Ön İşleme

- Eksik Verilerin İncelenmesi:**
Veride eksik değer olup olmadığı analiz edildi. Eksik veriler varsa uygun şekilde dolduruldu veya ilgili satırlar/sütunlar çıkarıldı.
- Aykırı Değer Tespiti ve Bayraklama:**
Özellikle sayısal değişkenlerde aykırı değerler (outliers) IQR yöntemi ile belirlendi. %10'dan fazla aykırı içeren sütunlara özel bayrak sütunları eklendi. Bu, modelin uç değerlerden daha az etkilenmesini sağladı.
- Çarpıklık (Skewness) Kontrolü ve Log Dönüşüm:**
Dağılımı çok sağa ya da sola yatık olan sütunlar ($skew > 1$ veya < -1) belirlendi. Pozitif ve çarpık değerler için log dönüşümü düşünülerek normal dağılıma yaklaştırılması hedeflendi.
- Düşük Varyanslı Sütunların Çıkarılması:**
Neredeyse sabit (çok düşük varyanslı) olan sütunlar belirlenerek modelden çıkarıldı. Bu tür sütunlar genellikle model performansına katkı sağlamaz.
- Korelasyon Analizi:**
Korelasyon matrisi oluşturularak birbirine çok benzeyen (yüksek korelasyonlu) sütunlar belirlendi. %90'dan fazla korelasyon içeren sütunlar arasından yalnızca temsilciler seçilerek bilgi tekrarının önüne geçildi.

Özellik Seçimi ve Veri Seti Hazırlığı

- Temsilci sütunlar, düşük korelasyonlu sütunlar ve hedef değişken (churn) birleştirilerek nihai veri seti oluşturuldu.
- Tüm sayısal değişkenler standartlaştırıldı (ölçeklendi) ve hedef değişken ayrı bir değişkene alındı.

Modelleme Süreci

- Veri Setinin Bölünmesi:**
Veri seti eğitim ve test seti olmak üzere %70-%30 oranında bölündü. Sınıf dengesizliğini korumak adına stratify parametresi kullanıldı.
- Model Eğitimleri:**

- **Random Forest:** Temel ağaç tabanlı model olarak kullanıldı. ROC AUC ve doğruluk skorları hesaplandı.
- **XGBoost:** Sınıf dengesizliğini dikkate alarak `scale_pos_weight` parametresiyle optimize edildi.
- **AdaBoost & Gradient Boosting:** Alternatif topluluk modelleri denenerek performans karşılaştırmaları yapıldı.

Model Değerlendirme ve Yorumlama

- **ROC AUC, Accuracy, Precision, Recall** gibi metrikler kullanılarak modeller karşılaştırıldı.
- **SHAP analizi** ile XGBoost modelinin hangi özelliklere göre karar verdiği yorumlandı.
- Ayrıca, **özellik önem sıralaması** çıkarılarak hangi değişkenlerin churn tahmininde en belirleyici olduğu görselleştirildi.

Sonuç ve Öneriler

Bu proje ile churn riskindeki müşteriler önceden tespit edilerek; özel kampanyalar, iletişim stratejileri ya da sadakat programları geliştirilebilir. Özellikle "gprs kullanımı", "son 100 çağrı süresi", "SMS harcamaları" gibi bazı davranışsal veriler churn üzerinde önemli etkiye sahiptir. Model sonuçlarına göre bu müşterilere özel eylem planları hazırlanmalıdır.