CHURN

telekom churn datası ile yapılan çalışmada öncelikle problemin tanımı, veri seti hikayesi ve parametrelerin neyi fade ettiği araştırıldı ve readMe file içerisine eklendi. Yapılan incelemede datanın tamamının 2013 yılına ait olduğu tespit edildi ve year parametresi bir anlam ifade etmediği için datasetten kaldırıldı. User\_account parametresi de her bir kullanıcının kimliğini ifade ettiği göz önünde bulundurularak analiz kapsamında bir işe yaramayacağı değerlendirilerek datasetten kaldırıldı. User\_lifetime değişkeni kullanıcının sistemde kalma süresini belirttiği için 0 olan değerler kaldırıldı. EDA.py ve data\_prep.py dosyaları dosyaları oluşturuldu. Bunun amacı oluşturulacak scriptlerin karmaşıklığını önlemek ve okuması kolay bir scrip oluşturmaktı. Exploratory\_data \_analysis.pynb dosyası colab platformunda oluşturuldu. İlk olarak datanın değişken tipleri, ilk ve son satırlarından örnekler, eksik veri olup olmadığı ve quantileları incelendi. Describe methodu yardımı ile de değişkenlerin mean, max, min gibi değerleri incelendi. Parametrelerin kategorik olanları ve numerik olanları tespit edildi ve buna göre her bir parametrenin hedef değişken ile olan ilişkisi incelendi. İnternet kullanımı, sms kullanımı ve sesli konuşma kullanımı gruplandırılarak inceleme kolaylaştırıldı. Churn\_analysis\_pipeline.pynb dosyasında ise kayıp veri tespiti, outlier tespiti yapıldı ve datasetten kaldırıldı. Kategorik olan değişkenlerin bir çoğu 0 ve 1den oluştuğu için encode etmeye gerek görülmedi. Mounth parametresi ayları belirttiği için one\_hot\_encoding yardımı ile encode edildi. Test ve train datası ayrıldıktan sonra model için 5 aday pretrained model seçildi ve base model olarak başarıları sınandı. Bu süreçte hem overfiti engellemek hem robustnesı sağlamak adına k katlı çapraz doğrulama kullanıldı. Başarılar incelendikten sonra hiper parametre optimizasyonu yapıldı. 3 farklı model seçilerek voting classifier oluşturuldu. Voting classifier oluşturmanın sebebi ise bir modele bağlı kalmadan 3 başarılı modelin kararı ile sonucu belirlemektir. Eğitim gerçekleştikten sonra başarısı incelendi. Başarıyı artırmak amacıyla feature engineerin işlemi gerçekleştirildi ve 5 yeni feature üretildi. Tekrar model eğiterek elde edilen featureların başarıyı artırıp artırmadığı incelendi ve artırdığı tespit edildi. Oluştutulan scriptler ve dosyalar github ortamına akatarıldı. Model boyutundan dolayı githuba yüklenemedi. Model validasyonu yapıldı sonuçlar scriptte mevcuttur.