

GSM Sistemlerinde Dolandırıcılık Risk Yönetimi

Onur TÜFEKÇİOĞLU
Prof. Dr. Banu DİRİ

Amaç

- GSM Operatörleri dolandırıcılık yapan aboneleri tespit etmek amacıyla çok ciddi yatırımlar yapmaktadır.
- Yapılan bu yatırımların karşılığında GSM operatörlerinin elde ettiği başarılar istenilen düzeyde olmamaktadır.
- Gerek abone memnuniyetini arttırma, gerekse yapılan yatırımların karşılığını alabilmek için daha isabetli bir dolandırıcılık tespit sistemine ihtiyaç vardır.

AMAÇ

- Daha iyi tahmin yapabilen bir dolandırıcılık tespit sistemi oluşturabilmek amacıyla 4 farklı makine öğrenmesi yöntemi ve derin öğrenme yöntemi ile mevcut sistem tarafından sınıflandırılmış abone kümesi (76 ve 10 özellik için) tekrar sınıflandırılmış ve her metodun başarısı ölçülmüştür.
- Hatalı aksiyon alınan abonelerle, hiç aksiyon alınmamış abonelerin başka operatörlere geçiş oranları karşılaştırılarak yapılan iyileştirmelerin olumlu etkileri ortaya konmuştur.

VERİ SETİ

- 1 Ağustos 2014 - 1 Ekim 2014 tarihleri arasında X GSM operatörüne ait dolandırıcılık şüphesiyle aksiyon alınan 5641 abonenin, dolandırıcılık durumları incelenmiş ve doğru tespit edilen dolandırıcılık vakalarının yüzdesini arttırabilecek 4 makine öğrenmesi yöntemi ile Derin Öğrenme yöntemi incelenmiştir.
- Aksiyon alınan abonelerin ses ve mesaj olmak üzere 2 farklı kullanım tipi incelenmiştir.
- Her kullanım tipi için servis ve tam kapama olmak üzere 2 farklı aksiyon tipi incelenmiştir.

VERİ SETİ

- Servis Kapama: Kullanım tipine göre SMS gönderimine yada Ses Kullanımına kapatma.
- Tam Kapama: Tüm kullanımlara kapatma

	servis	tam	Toplam
ses	3.743	1.019	4.762
mesaj	348	531	879
Toplam	4.091	1.550	5.641

VERİ SETİ

- Aksiyon alınan abone kitlesi aday ve dolandırıcı olarak sınıflandırılmıştır.
- Aday: Servis yada tam kapama aksiyonu alınan abonelerden kendisinden istenen ara ödeme tutarını ödeyerek hattını açtıran aboneler.
- Dolandırıcı: Servis yada tam kapama aksiyonu alınan abonelerden kendisinden istenen ara ödeme tutarını ödemeyenler.

VERİ SETİ

- İncelenen veri setinin aksiyon tipi ve kullanım tipine göre sınıflandırılması

	aday	dolandırıcı
ses	3.392	1.370
mesaj	572	307
ses+mesaj	3.964	1.677
servis	2.954	1.137
tam	1.010	540
servis+tam	3.964	1.677

VERİ SETİ

- Hazırlanan veri setinde toplamda 76 özellik vardır. Deneysel sonuçlar alınırken özellik seçimi yapılarak özellik sayısı 10'a indirilmiştir.
- Seçilen bu 10 özellik sırası ile
- *Abonelik_yasi* : *X* operatöründe aktif olduğu gün sayısı
- *Club* : Abonenin vip olarak kabul edilip edilmediği
- *Payorder* : Banka üzerinden otomatik fatura ödeme talimatının olup olmadığı
- *International_status* : Numaranın yurtdışı kullanıma açık olup olmadığı

VERİ SETİ

- *Csi_odeme_adet* : Operatördeki kayıtlı kredi kartı üzerinden yaptığı ödeme sayısı
- *Csi_toplam_odeme* : Operatördeki kayıtlı kredi kartı üzerinden yaptığı toplam ödeme tutarı
- *Ucretli_sms_cdr_adet_5tl* : 5 TL üzerindeki sms kullanım adeti
- *Toplam_arama_sayisi* : Toplam arama sayısı
- *Toplam_arama_suresi_50tl_uzeri* : Tutarı 50TL ve üzeri olan konuşmaların toplam süresi
- *Sayi_sn_5krs_uzeri* : Saniyesi 5 kuruş ve üzeri olan arama sayısı

İNCELENEN SİSTEMİN BAŞARISI

- İncelenen Sistemin alınan aksiyon tipi ve kullanım bazlı dolandırıcılık tespit etme başarısı aşağıdaki gibidir.

	servis	tam	servis+tam
ses	27,6	32,8	28,7
mesaj	29,0	38,8	34,9
ses+mesaj	27,7	34,8	29,7

- Mevcut sistemin ortalama başarısı %29,7'dir

İNCELENEN SİSTEMDE ABONE KAYIPLARI

- İncelenen Sistemin , hiç aksiyon alınmamış ve aksiyon alınmış abone bazında kayıp oranları (Başka operatörlere geçme)

	#abone	#kayıp	başarı %	katı
normal	38.988	36	0,0923	1
aday	3.964	179	4,51	49
dolandırıcı	1.677	95	5,66	61
aday+ dolandırıcı	5.641	274	4,85	52

- Aksiyon alınan aboneler ,hiç aksiyon alınmayan abonelere oranla yaklaşık 50 kat daha fazla başka operatörlere geçmektedir.

KULLANILAN YÖNTEMLER

Kullanılan makine öğrenmesi yöntemleri:

- k-En Yakın Komşuluk (k-NN)
- Naive Bayes (NB),
- Rasgele Orman (RF)
- Destek Vektör Makinesi (SVM)

KULLANILAN YÖNTEMLER

- Özellik seçimi ve abonelerin etiketlendirilmesinde açık kaynak kodlu Weka uygulaması kullanılmıştır
- Özellik azaltma ve modellenen sistemin başarısını ölçümülerken 10 katlı çapraz geçerleme (CV: Cross Validation $k=10$) kullanılmıştır.
- Özellik seçimi için de attribute evaluator: CfsSubsetEval, search method: BestFirst-D1-N5 kullanılmıştır

KULLANILAN YÖNTEMLER

Derin Öğrenme:

- Öğrenme algoritmalarını kullanabilmek amacıyla Matlab 2017 üzerinde yer alan hazır derin öğrenme kütüphanesi kullanılmıştır.
- Veri seti önce Autoencoder'dan geçirilmiş daha sonra, Softmax algoritması ve Destek Vektör Makinesi (DVM) ile 10 katlı Çapraz Geçerleme (CV: Cross Validation k:10) kullanılarak sınıflandırılmıştır.

SONUÇLAR

- Aksiyon Tipine Göre Başarı Oranları (%) (76 Özellik)

	servis	tam	servis+tam
k-NN	82,03	62,83	76,15
NB	54,14	63,41	54,42
RF	79,78	68,64	76,67
SVM	87,11	69,87	81,59

- *Aksiyon tipine göre başarı (%) (10 özellik).*

	servis	tam	servis+tam
k-NN	82,74	61,67	77,45
NB	57,95	57,35	67,50
RF	86,94	64	80,69
SVM	86,67	68,77	80,64

SONUÇLAR

- *Kullanım tipine göre başarı (%) (76 özellik).*

	ses	mesaj	ses+mesaj
k-NN	79,73	63,70	76,15
NB	57,29	64,05	54,42
RF	78,24	68,03	76,67
SVM	85,15	66,89	81,59

- *Kullanım tipine göre başarı (%) (10 özellik).*

	ses	mesaj	ses+mesaj
k-NN	80	68,71	77,45
NB	76,81	68,71	67,50
RF	84,73	68,48	80,69
SVM	83,93	66,21	80,64

SONUÇLAR

- *Aksiyon tipine göre başarı (%) (76 özellik).*

	Servis Kapama	Tam Kapama	Tüm Kapamalar
AutoEncoder +Softmax	77,22	59,31	68,58
AutoEncoder+DVM	77,68	53,68	66,76

- *Kullanım tipine göre başarı (%) (76 özellik).*

	Ses Kullanımı	Mesaj Kullanımı	Tüm Kullanımlar
AutoEncoder +Softmax	73,74	54,48	65,58
AutoEncoder+DVM	72,36	45,21	66,76

KARŞILAŞTIRMA

	Servis Kapama	Tam Kapama	Tüm Kapamalar
AutoEncoder +Softmax	77,22	59,31	68,58
AutoEncoder+DVM	77,68	53,68	66,76
K-NN	82,03	62,83	76,15
NB	54,14	63,41	54,42
RO	79,78	68,64	76,67
DVM	87,11	69,87	81,59
X Operatörü	27,7	34,8	29,7

KARŞILAŞTIRMA

	Ses Kullanımı	Mesaj Kullanımı	Tüm Kullanımlar
AutoEncoder +Softmax	73,74	54,48	68,58
AutoEncoder+DVM	72,36	45,21	66,76
K-NN	79,03	63,70	76,15
NB	57,29	65,05	54,42
RO	78,74	68,03	76,67
DVM	85,15	66,89	81,59
X Operatörü	28,7	34,9	29,7

SONUÇ

- *Mevcut sistemin başarısı %30 iken Rasgele Orman (RF) ve Destek Vektör Makinesi SVM yöntemlerinde %80 in üzerine çıkmaktadır.*
- *%50 'lik bir abone kitlesine hatalı aksiyon alınmasının önüne geçilebildiği görülmüştür.*
- *Aylık 100 civarı abonenin operatör değiştirmesinin engellenebildiği görülmüştür.*
- *Aylık 1500 aboneye hatalı aksiyon alınmasının önüne geçilebildiği görülmüştür.*



TEŞEKKÜRLER