

Lojistik Regresyon ile Bilgisayar Ağlarında Anomali Tespiti

İdris Budak¹, Baha Şen², Mehmet Zahid Yıldırım³

¹ Karabük Üniversitesi, Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümü, Karabük

² Yıldırım Beyazıt Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Bilgisayar Müh. Bölümü Ankara

³ Karabük Üniversitesi, Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümü, Karabük
idrisbudak@karabuk.edu.tr, bsen@ybu.edu.tr, m.zahidyildirim@karabuk.edu.tr

Özet: Bilgi çağının en önemli unsurları olan bilginin üretilmesi, korunması ve erişilmesinde kullanılan bilgisayar ağlarının hız, güvenlik ve sürekliliği bugün hayatın vazgeçilmezleri arasında ilk sıralarda yerini almaktadır. Bu yüzden ağdaki düzensizliklerin zamanında tespit edilip önlemlerin alınması gerekmektedir. Bu çalışmadaki amaç bilgisayar ağlarındaki anomali tespitinde Binary Lojistik Regresyon tekniğinin uygulanabilirliğini incelemektir. Bu amaçla Saldırı Tespit Sistemleriyle ilgili çalışmalarda en sık kullanılan veri setlerinden olan KDD Cup'99 veri seti kullanılarak bir matematiksel model oluşturulup bu modelin uygunluğu test edilmiştir.

Anahtar Sözcükler: Saldırı Tespit Sistemleri, Bilgisayar Ağlarında anomali Tespiti, Lojistik Regresyon.

Anomaly Detection in Computer Networks with Logistic Regression

Abstract: The most important elements of the information age is generation, securing, and access to knowledge, so the location of the first rank in indispensable of life today is computer networks with high speed, security, and continuity. Therefore, measures should be taken timely when anomalies detected in the network. The purpose of this working is to detect if logistic regression is usable in anomaly detection for computer networks. For this purpose we used KDD Cup'99 data set for mathematical model and tested that model for convenience.

Keywords: Intrusion Detection Systems, Anomaly Detection In Computer Networks, Logistic Regression.

1. Giriş

Yaşadığımız bilgi çağında şimdiye kadar hiç olmadığı kadar bilgi üretilmekte işlenmekte ve bunlara erişilmektedir. Bilginin bu denli hızlı üretilip yayılmasında hiç kuşkusuz bilgisayar teknolojileri en büyük teknik faydayı sağlamaktadır. Bilgiye erişim ve paylaşım için ise en fazla verimliliği bilgisayar ağları sağladığından, ağ işleyişinin düzgün olması hayati önem taşımaktadır. Ağ trafiğindeki anormallikler ise ağın gerektiği gibi kullanımını engelleyen unsurların başında gelmektedir. Bu anormallikler altyapı sorunlarından kaynaklanabileceği gibi ağın kötüye kullanılması veya ağa yapılan saldırılardan da kaynaklanabilmektedir.

Birçok kaynakta saldırı tespit sistemleri olarak da anılan anomali tespit sistemleri ağda oluşan düzensizlikleri tespit edip ilgili kişileri veya yazılımları uyarmayı sağlayan sistemlerdir.

Günümüzde çeşitli organizasyon ve kurumlar tarafından üretilmiş gerek ticari gerekse açık kaynak kodlu birçok saldırı tespit sistemleri mevcuttur. Bizim çalışmamızın özgün tarafı ise Logistic regresyonun çözüm yöntemlerinden olan logit modelin kullanılarak binary logistic regresyon ile tüm ağ trafiğinin analiz edilip belli bir anda trafikte anomali olma olasılığının ne olduğunu gösteren bir çalışma olmasıdır.

2. Anomali Tespitinde Kullanılan Yöntemler

Anomali tespit sistemleri, daha çok firewall'larda bulunan kural veya imza tabanlı sistemlerden farklı

olarak daha dinamiktir, ve henüz hakkında bir imza bilinmeyen saldırıları da algılama avantajına sahiptir.

Anomali tespitinde günümüze kadar en fazla istatistiksel yöntemler kullanılmasına rağmen bunun dışında: durum geçiş diyagramları (state transition diagrams), yapay sinir ağları (artificial neural networks), veri madenciliği (data mining), yapay bağışıklık sistemi (artificial immune system), örüntü eşleme, bulanık mantık (fuzzy logic) gibi farklı birçok yaklaşım uygulanmıştır.

3. Lojistik Regresyon

Lojistik regresyon analizinin kullanım amacı istatistikte kullanılan diğer model yapılandırma teknikleriyle aynıdır. En az değişkeni kullanarak en iyi uyuma sahip olacak şekilde sonuç değişkeni (bağımlı yada cevap değişkeni) ile bağımsız değişkenler kümesi (açıklayıcı değişkenler) arasındaki ilişkiyi tanımlayabilen ve genel olarak kabul edilebilir modeli kurmak. [1]

Lojistik regresyon analizi sonucunda elde edilen modelin uygun olup olmadığı “model ki-kare” testi ile, her bir bağımsız değişkenin modelde varlığının anlamlı olup olmadığı ise Wald istatistiği ile test edilir.

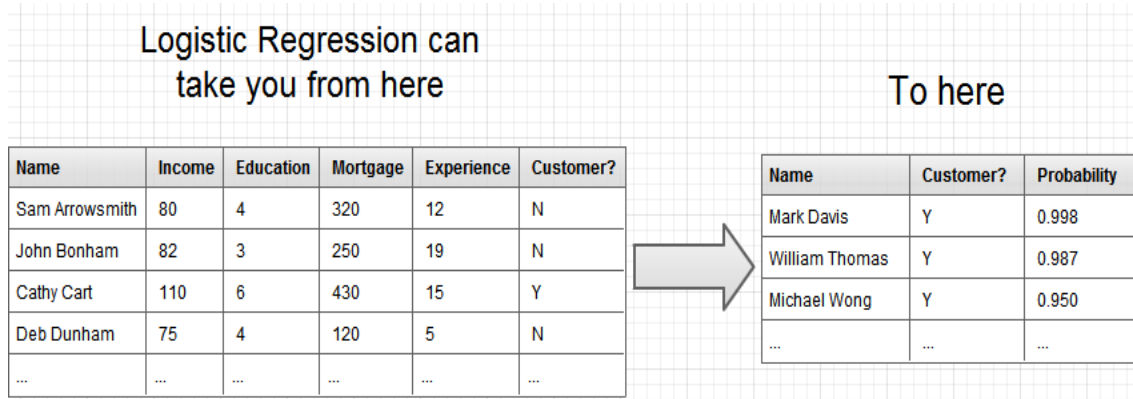
Bazı nicel değişkenleri regresyon modellerinde nitel olarak kullanmak da mümkündür. Örneğin sınav sonuçları sayısal not olarak alınabileceği gibi kötü-iyi-çok iyi gibi 3 değerli nitelik olarak da alınabilir.

3.1 Lojistik Regresyon ile Doğrusal Regresyon Farkı:

Lojistik Regresyon ile Doğrusal Regresyonun en temel farkı doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyonda ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı kestirilir. (Çoşkun v.d, 2004:43)

Bilinen doğrusal regresyon analizinde bağımlı değişken ve bağımsız değişken(ler) sayısal (ölçümle belirtilen sürekli ya da kesikli sayısal) olarak belirtilir. Örneğin, yaş ile kan basıncı arasında bir ilişki aranacaksa; hem yaş, hem de kan basıncı sayısal olarak belirtilmelidir. Nitelik olarak belirtilemezler.

Bağımlı değişken nitelik olarak belirtilirse, bağımsız değişken ya da değişkenlerle arasındaki ilişki lojistik regresyon yöntemiyle aranır. [2]



Şekil [3]

4. Lojistik Regresyon Modelleri

Log-linear, Logit ve Probit Modeller iki şıklı bağımlı değişkenleri açıklamada regresyon gibi genel doğrusal modellerin temelini oluşturmaktadır. Bu modeller bağlantı fonksiyonu olarak Sıradan En Küçük Kareler tahmini yerine Maksimum Benzerlik (En Çok Olabilirlik) tahminini kullanarak standart regresyondan ayrılır. [4]

Logistic Regresyon genel olarak üçe ayrılır:

- 1- İkili (Binary) lojistik regresyon: Bağımlı değişken iki düzeyli olduğunda kullanılır(Var-Yok, Evet-Hayır).
- 2- Sıralı (Ordinal) lojistik regresyon: Bağımlı değişken sıralı nitel veri tipinde (hafif-orta-şiddetli vb.) olduğunda kullanılır.
- 3- Multinomial lojistik regresyon: Bağımlı değişken ikiden çok düzeyli sıralı olmayan nitel veri tipinde olduğunda kullanılır.

4.1 Logit Model:

Odds, başarı ya da görülme olasılığının “p”, başarısızlık ya da görülme olasılığına “1-p” oranıdır.

Odds ratio (OR) ise iki odds’un birbirine oranıdır. İki değişken arasındaki ilişkinin özet bir ölçüsüdür. (Bahis Oranı da denir)

Tablo1: UDP Protokolü ODDS Değerleri

Protokol udp mi?	Saldırı		Toplam
	Var	Yok	
Evet	3	127	130
Hayır	45	180	225
Toplam	48	307	355

Örneğin yukardaki Tablo1’e baktığımızda:

Protokol tipi udp olan bağlantılarda saldırı olma odds’u:

$$(3/130) / (127/130) = 3/127 = 0.024$$

Udp olmayan bağlantılarda ise:

$$45/180 = 0.25$$

$$\text{Odds ratio} = 0.25 / 0.024 = 10.4$$

Bu bize udp protokolü kullanmayan bağlantıların saldırı olma olasılığının, udp kullananlardan yaklaşık 10 kat daha fazla olduğunu göstermektedir.

Logit ismi, odd değerinin doğal logaritmasını ifade etmektedir. Yani π olasılığı göstermek üzere, logit;

$$\text{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right)$$

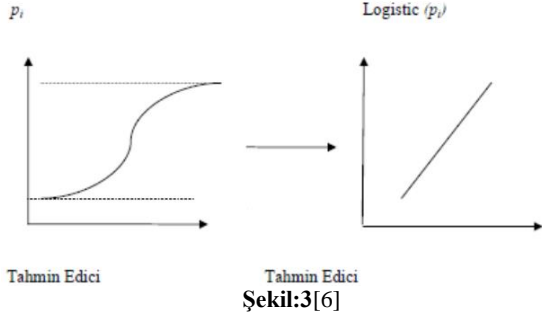
Şekil:2

Logit model, bağımsız değişken değeri sonsuza gittiği zaman, bağımlı değişkenin 1’e asimptot olduğu matematiksel bir fonksiyondur.[5]

Logit modellerinde olasılıklar 0 ile 1 arasında sınırlandırılmışlardır. Bunu yaparken lojistik regresyon

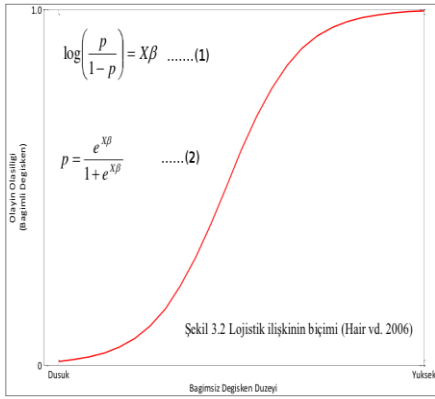
modeli olasılıklara bir dönüşüm uygulamaktadır; çünkü olasılıklar ve tahmin edici değişken arasındaki ilişki doğrusal değildir ve S şeklinde bir eğridir.

Aşağıda Şekil3'ten görüleceği üzere, lojistik regresyon varsayımı altında olasılıkların lojistik dönüşümü ok ile gösterilmekte olup bu dönüşüm, tahmin edici değişkenler ile olasılıkların doğrusal bir ilişki içerisinde sonuçlanmasını sağlamaktadır. [6]



4.2 Lojistik Regresyon Formülü:

Lojistik regresyon fonksiyonu bağımlı ve bağımsız değişkenler arasında aşağıdaki lojistik fonksiyonunu kullanmaktadır:



Şekil:4[7]

Yukardaki Şekil4'te geçen formülde p olayın olma olasılığı (bağımlı değişkenin tahmin edilen değerini) vektörünü, β model parametreleri vektörünü, X ise sabit terimi de içerisinde barındıran bağımsız değişkenler matrisini temsil etmektedir. (p) vektörü aşağıdaki formülle hesaplanmaktadır.[7]

$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}}$$

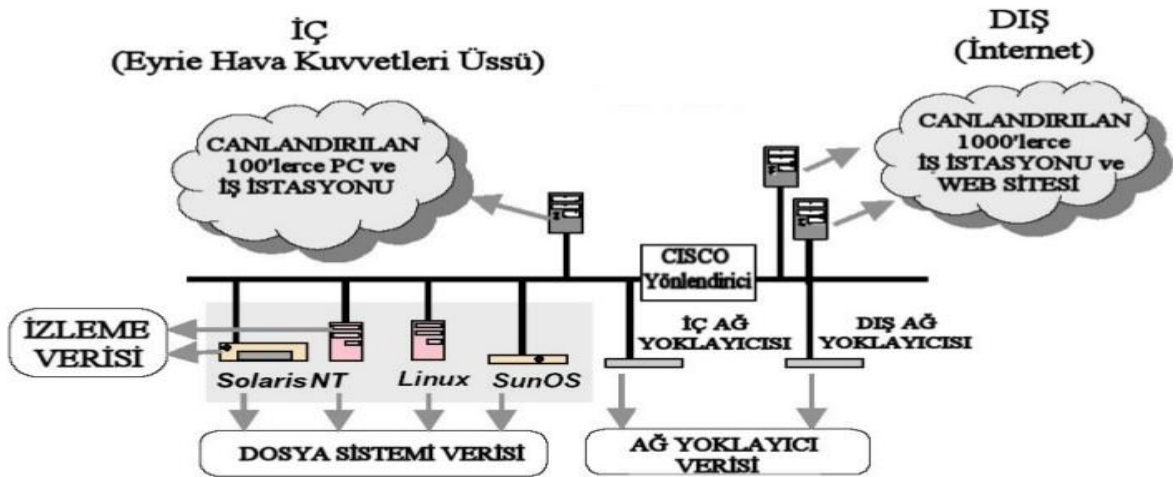
Şekil:5

4.3 Logit Model ile Lojistik Regresyon Arasındaki Benzerlik ve Farklar

Lojistik regresyon, bir ya da daha fazla açıklayıcı değişken ile ikili bir yanıt değişkeni arasındaki ilişkiyi göstermek için kullanılan bir tekniktir. Açıklayıcı değişkenler kategorik, sürekli ya da her ikisi de olabilir. Logit modellerde ise açıklayıcı değişkenler sadece kategorik değişkenlerden oluşabilir.[8]

5. Kullandığımız Veri Seti

Saldırı Tespit Sistemleriyle ilgili çalışmalarda en sık kullanılan veri seti DARPA 1998 ve 1999 veri setleridir. Biz de model oluşturma çalışmamızda yine bu verilerden türetilen KDD Cup'99 veri setini kullanacağız. Veri setini oluşturan kaynak aşağıdaki Şekil6'da da görüldüğü gibi saldırının hedefi olan bir iç ağ ve saldırıyı gerçekleştiren bir dış ağ olmak üzere iki farklı ağdan oluşmaktadır



Protokollerin (HTTP, SMTP, telnet,...) karışımı, trafik yoğunluğunun saatlik değişimleri, 1998’de gerçek Hava Kuvvetleri ağından toplanan trafiğe benzer olacak şekilde tasarlanmıştır. Ağ üzerinden 2 noktadan veri toplanmıştır: dört kurban makine ile yönlendirici arasındaki iç ağ dinleyicisi ve yönlendirici ile İnternet arasındaki dış ağ dinleyicisi üzerinden. [9]

DARPA verileri ile çalışırken matlab ya da sql sunucularla birlikte tcpdump çıktılarını wireshark(Eski adı ethereal) programıyla da inceleyebiliriz. Ağ dinleyicisi iki yönlü paketleri yakaladığı için kurban makinalara gelen paketler için varış ip adresi 172.16.x.x olan paketler olarak süzülmalıdır. Örnek bir filtre:

```
(ip.dst == 172.16.0.0/16) and !(ip.src == 172.16.0.0/16) and !(ntp) and !(rip) and !(loop) and !(arp) and !(nbns) [10]
```

5.1 Veri Setinin Hazırlanması:

Biz çalışmamızda “kddcup.data_10_percent_corrected” dosya ismi ile internette indirilebilen yaklaşık 75Mb büyüklüğünde ve içinde yaklaşık 500bin kayıt bulunan gerçek kddCupp-99 veri seti’nin 10%’una karşılık gelen veri setini kullandık. Veri setimizin ilk 250bin kaydını model oluşturmak için kalanı ise test için kullandık. Veri setinde toplam 41 adet değişken bulunmaktadır. Biz aşağıdaki prensiplere uyarak bu sayıyı 9’a indirdik:

- Paketlerin sadece başlık bilgisine değil içeriğine de bakılarak anlaşılacak alanlar da alınmıştır.(Örneğin bu yüzden hot, su_attempted gibi alanlar alınmıştır.)

- Parametrelerin birbirlerinden bağımsız olanları seçilmiştir. Örneğin root_shell, su_attempted, num_root alanlarının tümü birden alınmak yerine su_attempted alanı alınmıştır.

- Parametrelerin bağımlı değişkeni etkilemeyecek olanları seçilmemiştir. Örneğin src_bytes ve dst_bytes alanları bu yüzden alınmamıştır.

Verileri incelemek ve binary hale getirmek için öncelikle verileri sql sunucusuna alıp aşağıdaki kurallara uygun olarak ikili hale getirdik:

protocol_type: tcp=1 ; udp veya icmp=0.

Bağlantının(connection, datasetteki her satır bir bağlantıdır.)

service: smtp, ftp, pop_3, ldap, login, imap4, auth, IRC, telnet, sql_net, exec, shell, klogin, kshell = 1 , diğerleri = 0 .

Hedefteki ağ servisini gösterir.

flag: SF veya OTH = 0 ; diğerleri 1.

SF bağlantının normal bir şekilde sonlandığını, OTH ise bağlantı takip işinin bağlantının ortasında başladığını gösterir.

land: Hedef ve kaynak ip/port bilgileri aynı ise = 1 ; değilse 0.

wrong_fragment: sıfır ise=0; değilse=1.

Hatalı fragment sayısını gösterir.

hot: sıfırdan büyük ise 1 ; değilse 0 .

Bir bağlantıda çalıştırılan kritik komut sayısını

gösterir. Örneğin sistem klasörüne girmek, programlar oluşturup çalıştırmak gibi.

num_failed_logins: sıfırda 0 ; değilse 1 .

Yanlış login işlemleri sayısını gösterir.

su_attempted: “su root” komutu denenmişse 1 diğer durumda 0.

num_access_files: sıfırdan büyük ise 1 ; değilse 0 .

Kontrol ya da erişim izinlerini tutan kritik dosyalarda yapılan işlem sayısı.

Verilerimizin örnek görüntüsü aşağıda Tablo2’deki gibi oldu: (En son kolondaki label bu kaydın-satırın-saldırı olup olmadığını tutmaktadır, saldırılar için 1 , normal kayıtlar için 0 değerini verdik.)

Tablo2: Veri Seti Örnek Görüntüsü

protocol_type	service	flag	land	wrong_fragment	hot	num_failed_logins	su_attempted	num_access_files	label
1	1	0	0	1	1	1	0	1	1
0	0	1	1	1	0	0	0	1	0
1	0	0	1	0	1	0	1	1	0

5.2 Modelin Oluşturulması:

SPSS yazılımını kullanarak veri setimizi binary logistic regresyon ile analiz ettiğimizde aşağıdaki sonuçları elde ettik:

Tablo3: Durum İşleme Özeti

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	250000	100,0
	Missing Cases	0	,0
	Total	250000	100,0
Unselected Cases		0	,0
Total		250000	100,0

Yukardaki Tablo3 veri setimizdeki tüm kayıtların analiz için kullanıldığını göstermektedir. Veri setimizde 250bin kayıt vardı, tabloda görüldüğü üzere tümü incelemeye alınmış.

Tablo4: Bağımlı Değişken Kodlaması

Original	Internal Value
0	0
1	1

Tablo4 SPSS'in, bağımlı değişkenimiz için, veri setimizdeki "0"ları binary "0" olarak, "1"leri ise binary "1" olarak aldığını göstermektedir, ki biz de böyle olması için data setimizi önceden buna göre hazırlamıştık. Çünkü bizim incelemek istediğimiz konu saldırı olma durumu olduğundan saldırı olması durumunu 1, olmaması durumunu ise 0 olarak kodlamıştık. Bu tamamen bir tercih meselesidir ve tamamen tersi de seçilebilirdi. Fakat sonuçların yorumunun kolay olması için genelde asıl ilgilendiğimiz cevap için "1" kullanmamız işimizi daha kolaylaştırır.

"Block 1: Method = Enter"

SPSS çıktısındaki yukardaki ibare ise metod olarak girişi(Enter) seçtiğimizi yani adımsal(stepwise) ya da hiyerarşik metodları kullanmadığımızı gösterir.

Tablo5: Sınıflandırma Tablosu

Observed		Predicted			
		label		Percentage Correct	
		0	1		
Step 1	label	0	59805	11420	84
		1	736	178039	99,6
Overall Percentage					95,1
a. The cut value is 0,5					

Tablo6: Eşitlikteki Değişkenler

Step 1 ^a		B	S.E.	Wald	df	Sig.	Exp(B)
	protocol_type	-7,133	,039	3,380E4	1	,000	,001
	service	-1,372	,068	401,313	1	,000	,254
	Flag	6,882	,041	2,811E4	1	,000	974,877
	land	34,951	3,446E7	,000	1	1,000	1,509E15
	wrong_fragment	34,700	6,372E6	,000	1	1,000	1,175E15
	Hot	6,454	,071	8,375E3	1	,000	635,420
	num_failed_logins	3,803	,889	18,313	1	,000	44,854
	su_attempted	2,610	1,434	3,315	1	,069	13,602
	num_access_files	-,267	,486	,302	1	,583	,765
	Constant	2,843	,012	5,856E4	1	,000	17,175

a. 1. Adımda ele alınan değişkenler: protocol_type, service, flag, land, wrong_fragment, hot, num_failed_logins, su_attempted, num_access_files.

Yukardaki Tablo6 ise kullandığımız değişkenlerin hangilerinin hangi katsayılarla nihai modelimizde bulunacağını göstermektedir. Örneğin tabloda "service" değişkenimizin katsayısının(coefficient değerinin) -1,372 olması demek değişkenimizdeki 1 birimlik artışın(yani 0 yerine 1 olmasının başka bir deyişle "http" yerine "telnet" olmasının) sonucun log odds(yani logit) değerini -1,372 oranında düşürdüğünü gösterir.

"land" ve "wrong_fragment" hariç diğer değişkenlerimiz içinde mutlak değer olarak en büyük katsayıya -7,133 ile "protocol_type"; en küçüğüne ise -0,267 ile "num_access_files" değişkenlerimizin sahip olduğunu görmekteyiz. Bu da bize bir kaydın saldırı olup olmadığını belirlemede en büyük belirleyici role sahip parametrenin "protocol_type" olduğunu; en az ağırlığın ise "num_access_files" değişkeninde olduğunu göstermektedir.

Parametrelerimizin S.E.(Standart Error) değeri ise tahminimizin ne kadar stabil olduğunun ölçüsüdür ve ne kadar düşüğe o kadar tutarlı sonuçlar alırız. Örneğin "land" ve "wrong_fragment" hariç diğer tüm değişkenlerimizin ortalaması yaklaşık 0.4 iken bu iki değişkenin ortalaması yaklaşık 5 civarındadır, bu da bu iki değişkenin stabiliteyi ne kadar fazla bozduğunu göstermektedir.

Wald istatistiği değişkenlerimizin hangilerinin modelimiz için anlamlı olduğunu hangilerinin gereksiz olduğunu göstermektedir. Değişkenlerimiz içinde wald değeri sıfıra çok yakın olan "land" ve "wrong_fragment" değişkenlerimizin modelde gereksiz olduğu sonucu çıkmaktadır.

Tablodaki "Sig." kolonu ise değişkenin anlamlılık düzeyini göstermekte olup SPSS'te varsayılan olarak p<0.05 olarak çalışmaktadır. Sig değeri sıfıra ne kadar çok yakınsa parametrenin modeldeki anlamlılık düzeyi o kadar fazla demektir. Buna göre Sig değeri 1 olan "land" ve "wrong_fragment" değişkenlerimizin modelimizde anlamlı olmadıkları sonucu çıkmaktadır. Sig, wald ve SE değerlerinden anlaşıldığı üzere "land" ve "wrong_fragment" değişkenlerimiz nihai modelimizde bulunmayacaklardır. Buna göre nihai modelimiz aşağıdaki gibi olacaktır:

Regresyon eşitliği aşağıdaki gibi olmak üzere:

$$g(x) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k$$

$$g(x) = 2,843 + \text{protocol_type}*(-7,133) + \text{service}*(-1,372) + \text{flag}*6,882 + \text{hot}*6,454 + \text{num_failed_logins}*3,803 + \text{su_attempted}*2,610 + \text{num_access_files}*(-0,267)$$

$$P = 1/(1 + e^{-g(x)})$$

5.3 Modelin Uygulaması:

Örneğin aşağıdaki gibi bir kayıt için modelimizin ürettiği sonuca bakalım:

protocol_type=tcp, service=telnet, flag=S0, hot=0, num_failed_logins=0, su_attempted=0, num_access_files=0, label=neptune.

Kaydın label yani saldırı olup olmadığı ile ilgili bilgi alacağımız alanında “neptune” yazmaktadır. Yani bu bir neptune saldırısıdır.

Her parametreyi iki kategorili hale çevirip $g(x)$ fonksiyonunda yerine koyarsak :

$$g(x) = 2,843 + 1*(-7,133) + 1*(-1,372) + 1*6,882 + 0*6,454 + 0*3,803 + 0*2,610 + 0*(-0,267) = 1,22$$

$$P = 1/(1 + e^{-g(x)}) = 1/(1 + e^{-1,22}) = 0.7721$$

yani sonuçta bu kaydın yaklaşık 77% ihtimalle saldırı olduğunu söyleyebiliriz. Bizim modelimizde eşik(cutoff) değerimiz 50% olduğundan ve $77 > 50$ olduğundan modelimiz sonuç için 100% doğrulukla bu bir saldırı demektir.

Aşağıdaki Tablo7 ise test için kullandığımız veri setimize(kddCupp-99 veri seti'nin 5%ini içeren yaklaşık 250bin kayıt) modelimizin uygulanması ile elde ettiğimiz sonuçları göstermektedir. Modelimiz test verileri üzerinde gerçekte saldırı olan kayıtları 99,99% oranında doğru saptayabilirken, gerçekte saldırı olmayan kayıtlarda ise 42% oranında doğru sonuç üretebilmektedir.

Tablo7: Test Verisi Sınıflandırma Tablosu

Gerçek	Tahminimiz			
	label	0	1	Doğruluk Yüzdesi
label	0	10912	15141	42
	1	11	217957	99,9
Toplam Yüzde				93,8

Tahminimiz $\geq 0,5$ ise saldırı kabul ettik.

5.4 Sonuç ve Öneriler:

Sonuç olarak diyebiliriz ki modelimizin analiz ettiği bir kayıt gerçekte bir saldırı ise bunu 99%un üzerinde bir ihtimalle saldırı olduğunu bulabiliyoruz. Ki bu oran çok yüksek bir başarı oranıdır. Fakat gerçekte saldırı olmayan bir kaydın analizinde modelimizin başarı oranı biraz düşük (modeli bulduğumuz veri setinde 84%. Test veri setimizde ise 42%). Ki bu da modelimizin sürekli saldırılara maruz kalan , fakat

güvenlik seviyesi çok yüksek olması gereken, hiçbir saldırıya tahammülü olmayan, yanlış alarmlarla (false-pozitif) uğraşacak yeterli elemanı olan, kritik öneme sahip ağ işletim merkezleri için son derece uygun bir model olduğunu göstermektedir.

İlerki çalışmalarda günümüz internet trafiği verilerinin özelliklerini de dikkate alıp bu modeli geliştirerek bunu kullanan bir yazılım üretip gerçek hayatta kullanılabilir. Biz bir sonraki çalışmamızda aynı veri seti ve parametreleri kullanarak yapay sinir ağları ile de bir model oluşturup iki modelin karşılaştırılmasını sağlayacağız.

6. Kaynaklar

[1] “Lojistik Regresyon Analizinin İncelenmesi Ve Diş Hekimliğinde Bir Uygulaması” Sibel COŞKUN , Doç. Dr. Mahmut KARTAL, Yrd. Doç. Dr. Akın COŞKUN, Yrd. Doç. Dr. Hüdaverdi BİRCAN

[2] Lojistik Regresyon Analizi http://78.189.53.61/bs/ess/k_sumbuloglu.pdf Sayfa Görüntüleme Tarihi: 10.12.2012

[3] Bala Deshpande "Understand 3 critical steps in developing logistic regression models " "<http://www.simafore.com/blog/bid/99443/Understand-3-critical-steps-in-developing-logistic-regression-models>" Sayfa Görüntüleme Tarihi: 10.12.2012

[4] Yapay Bağımlı Değişkenli Tahmin Modelleri Ve Bir Uygulama, “Tuğba Altıntaş” , Yüksek Lisans Tezi , “İstatistik Anabilim Dalı”

[5] "Doğrusal Olasılık ve Logit Modelleri ile Parametre Tahmini" "M. Emin İnal" , "Derviş Topuz" , "Okyay Uçan"

[6] Dr. Göknur Büyükkara "http://www.acikders.org.tr/pluginfile.php/3496/mod_resource/content/2/Kredi_Riski.pdf" Sayfa Görüntüleme Tarihi: 10.12.2012

[7] Yemeklik Yağ Sektöründe Tüketici Davranışlarını Etkileyen Faktörlerin Analizi “Dr. Flora POLAT”

[8] Multinomial Logit Modeller Ve Bir Uygulama. Sevilay Karahan "Biyostatistik Programı" Yüksek Lisans Tezi

[9] M. A. Aydın, “Bilgisayar Ağlarında Saldırı Tespiti için İstatistiksel Yöntem Kullanılması”, İTÜ Yüksek Lisans Tezi, 2005.

[10] Saldırı Tespit Sistemlerinde İstatistiksel Anormallik Belirleme Kullanımı "Bahar 2005" Yük. Müh. Melike Erol