Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 29, S. 122-128, Aralık 2021

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 29, pp. 122-128, December 2021

Research Article

Kötücül URL Filtreleme için Derin Öğrenme Modeli Tasarımı

Recep Sinan Arslan[1*]

[1*] Kayseri Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye (ORCID: 0000-0002-3028-0416)
sinanarslanemail@gmail.com

(International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2021 – 21-23 October 2021)

(**DOI:** 10.31590/ejosat.1011961)

ATIF/REFERENCE: Arslan, R.S. (2021). Kötücül URL Filtreleme için Derin Öğrenme Modeli Tasarımı. *Avrupa Bilim ve Teknoloji Dergisi*, (29), 122-128.

Öz

Web saldırılarında yeni tekniklerin kullanımı ile birlikte birçok web uygulaması çeşitli güvenlik tehditlerine ve ağ saldırılarına maruz kalmaktadır. URL adresleri de bu güvenlik mimarisinin odak noktasını oluşturmaktadır. Birçok web uygulamasına URL adresleri üzerinden erişim sağlanmaktadır. Bu durum siber korsanların, URL adreslerini suç işlemek için kullanabilecekleri bir araç haline getirmektedir. Son kullanıcıları korumak amacıyla bu adreslerin tespit edilerek nasıl filtreleneceği çözülmesi gereken bir problemdir. Bu çalışmada kötü amaçlı URL adreslerinin tespiti için derin öğrenme ağı(DNN) tasarlanmıştır. Çalışmanın ilk aşamasında URL adresleri metin tabanlı analiz yapılarak işlenmiştir. Sonrasında 1 giriş, 3 gizli ve 1 çıkış katmanından oluşan DNN modeli sınıflandırma için eğitilmiştir. Model ISCX-URL2016 veriseti ile test edilmiş olup deneysel sonuçlar önerilen yapının yüksek hassasiyetli sınıflandırma için uygun olduğunu göstermiştir. Verisetinde iyicil 7781, tahrif edilmiş 7930, kimlik avı 7586, kötü amaçlı yazılım dağıtan 6712 ve spam türünde 6698 örnek bulunmaktadır. Her bir örnek için 79 özellik bulunmaktadır. Deneyler sonucunda 5 sınıftan oluşan problem için %95,4 doğruluk, %95,5 kesinlik, %95,4 duyarlılık ve f skoru değerine ulaşılmıştır. Bu çalışmanın birinci aşamasında Doc2Vec ağı kullanılarak özellikler çıkarılmıştır. Doc2Vec kullanılarak yapılan sınıflandırmada çok sınıflı problem için alınan %88.1 doğruluk değeri, bu çalışmada %95,4'e yükseltilmiştir. Metin tabanlı analizin vektör tabanlı analize göre çoklu sınıflandırma için daha başarılı olduğu gösterilmiştir. Sonuçta, web sitelerini ziyaret edenlerin niyetlerini belirlemek için URL adreslerini kullanmak etkin bir yöntemdir. Derin öğrenme modellerinin kullanılması web araştırmaları için önemli teorik ve bilimsel değere sahiptir ve güvenlik internet ortamı için farklı imkânlar sağlamaktadır.

Anahtar Kelimeler: Kötücül URL Filtreleme, Ağ Güvenliği, Derin Öğrenme, Web Atakları

A Deep Learning Model for Malicious Url Filtering

Abstract

Many web applications are exposed to various security threats and network attacks with the use of new techniques in web attacks. Url addresses are also the focus of this security architecture. Many web applications are accesses via Url addresses. This makes it a tool that hackers can use to commit crimes. In order to protect the end users, how to detect and filter these addresses is a problem that needs to be solved. In this study, a deep neural network (DNN) is designed for the detection of malicious Urls. In the first stage of the study, Url addresses were processed by making text-based analysis. Afterwards, the DNN model consisting of 1 input, 3 hidden and 1 output layers is trained for classification. The model was tested with the ISCX-URL2016 dataset and the experimental results showed that the proposed structure is suitable for high precision classification. The dataset includes 7781 benign, 7930 defacement, 7586 phishing, 6712 malware and 6698 spam urls. There are 79 features for each sample. As a result of the experiments, 95.4% accuracy, 95.5% precision, 95.4% sensitivity and f-score values were achieved for the problem consisting of 5 classes. In the first stage of this study, features were extracted using the Doc2vec network. In the classification made using Doc2vec, the accuracy value of 88.1% for the multi-class problem was increated to 95,4% in this study. It has been shown that text-based analysis is more successful for multiclass classification then vector-based analysis. After all, using Url addresses is an effective method to determine the intentions of website visitors. The use of deep learning models has important theoretical and scientific value for web research and provides different possibilities for the security internet environment.

Keywords: Malicious Url filtering, network security, deep learning, web attacks

[*] Sorumlu Yazar: sinanarslanemail@gmail.com

1. Giriş

İnternet teknolojilerinin gelişmesi ve kullanımının yaygınlaşması ile birlikte siber saldırganlar giderek daha önemli bir güvenlik sorunu haline gelmişlerdir. Kimlik avı, truva atları gibi birçok kötücül yazılım türü saldırı amacı ile internet adreslerini yani URL adreslerini bir araç olarak kullanmaktadırlar. URL adreslerinin üretimi ile ilgili algoritmaların belirli seviyelere ulaşması nedeniyle hergün yeni ve çok sayıda kötücül URL adresi ortaya çıkmaktadır. Bu nedenle çeşitli ağ saldırılarını önlemek ve ağ güvenliğini sağlamak için bu web linklerinin belirlenmesi oldukça önemlidir (He ve ark., 2021). Kaspersky güvenlik istatistiklerine göre tespit edilen tehditlerin %85.40'ı kötücül URL adresleridir (Malware variety grows, 2019).

Siber korsanlar genellikle hedefledikleri web sayfasına benzer bir web sitesi oluştururlar veya tarayıcı tarafında bulunan güvenlik açıklarından yararlanabilecek kod parçacıklarını yerleştirirler. Hedefledikleri kullanıcıların bilgilerini elde etmek veya bilgisayarlarını kontrol etmek için bu sayfalara yönlenmesini sağlamaya çalışırlar. Bu zamana kadar bu sitelerin tespit edilmesine yönelik olarak birçok çalışma yapılmış olup, bu çalışmaların birçoğu web sitesine ait özelliklere dayanmaktadır (Yuan ve ark., 2021). Web sitesine ait özellikler; URL-tabanlı özellikler, host tabanlı özellikler ve içerik tabanlı özellikler olmak üzere 3'e ayrılmaktadır. URL tabanlı özellikler esas olarak host, yol ve parametrelerden çıkarılan özellikleri kullanır. Her bir bölümde, kullanılan kelimeler ile temsil edilen kaynak arasında mantıksal ilişkiyi aranır. Devi(Devi ve ark., 2009)'nin araştırması göstermiştir ki, URL adresleri web siteleri hakkında zengin bilgiler taşır ve sınıflandırma için kullanılabilirler. Bunun yanında domain adresleri birden çok sözcük ve ayırıcılardan(nokta, /, ?) oluşur ve bir ip adresini temsil eder. Normal bir domain adresi genellikle ait olduğu kuruluşu temsil eder. Palaniappan (Palaniappan ve ark., 2020) DNS üzerinden domain adreslerinin analizi ile kötücül web sitelerinin tespitini yapmıştır. Host tabanlı özellikler internet alan adları, ip adresleri, alan adı sahiplik bilgisi gibi özellikleri ifade eder. Kötü amaçlı web sitelerinin kullandığı alan adları sık sık değişebilir. McAfee kötü niyetli saldırıları tespit etmek için alan adı yaşına bakmaktadır(Ebeling, 2021). Genel olarak kötücül web sitelerinin alan adlarının hayatta kalma süreleri normal kuruluş alan adlarından daha kısadır. Stevanovic (Stevanoviz ve ark., 2015) kötü amaçlı web sayfalarını tespit etmek için DNS trafiğini incelemektedir. Böylece host tabanlı olarak elde ettiği özellikleri sınıflandırma içi kullanmıştır. İçerik tabanlı özellikler ise ilgili web sayfasının html, kod ve görsel elementlerinin incelenmesine dayalıdır. Html bir dizi etiket ve sayfanın kaynaklarını biçimlendirme için kullanılır. Hou ve ark. (Hou ve ark., 2010) daha iyi tespit yöntemi geliştirmek için form ve sözcüksel özelikleri kullanmışlardır. Web sayfası kod bölümünde Javascript daha özgün ve son kullanıcı ile etkileşimli bir sayfa oluşturmak için uygundur ancak saldırganlar tarafından saldırı için kullanılmaktadırlar. Huang ve ark. (Huang ve ark., 2021) JSCortana yönteminde uyarnabilir bağlam analizi ve verimli anahtar özellik çıkarımı yaklaşımı kullanmışlardır. Javascript kodunun ayrıntılı analizi yapılmakta ve web sayfalarının güvenliğine yönelik bir öneri ortaya koyulmuştur.

Son kullanıcılar, saldırganlar tarafından hazırlanan ve internet üzerinden yayınlanan kimlik avı web sitesi adreslerine tıkladıklarında saldırıya uğrarlar. Bu nedenle ilgili siteye yönlendirilmeden önce bunun kullanıcılara bildirilmesi gereklidir. URL tabanlı yöntem, sayfa üzerinden herhangi bir analiz ve ayrıştırma işlemi yapılmadığından, diğer yöntemlerden daha hızlıdır. Makine öğrenimi algoritmalarına dayalı URL sınıflandırmasına yönelik çalışmalarda özelliklerin çıkarılması ve seçimi karmaşık bir süreci ifade eder. Sonuçta ortaya çıkan özellik vektörleri metne bağlı bilgileri ve host bilgilerini içerir. Metin özellikleri aslında URL'e ait bir takım bilgiler içerse de çoğu sınıflandırma için ayırt edici bilgiler değildir. URL sınıflandırma için sinir ağı modelinin kullanıldığı birçok çalışma bulunmakta olup genel olarak başarılı sonuçlar elde edilmiştir. Bu çalışmanın birinci aşaması olarak yapılan çalışmamızda(Arslan, 2021) ISCXURL2016 veri setinde bulunan URL adreslerinden Doc2Vec derin öğrenme ağı ile özellikler çıkarıldı ve hem ikili hem de çoklu sınıflandırma için modeller önerilmiştir. Buna göre ikili sınıflandırma yüksek başarı düzeyi yakalanırken kötücül URL tiplerini de alt sınıflara ayıracak şekilde çok sınıflı problem için %88.1 gibi düşük değer elde edilmişti. Bu çalışmada ise çoklu sınıflandırma ortaya çıkan bu durumu çözümlemek üzere URL adreslerinin lexical özelliklerinden(sunucu adı, URL uzunluğu, URL adresinde bulunan bölümler vb.) üretilen 27 farklı değerin kullanıldığı bir özellik vektörü ile çalışılmıştır. Böylece çoklu sınıflandırma da yüksek başarı düzeyinin yakalanması amaçlanmıştır. Sonuçta Doc2vec ile elde edilen özellikler ikili sınıflandırmada daha başarılı iken, çoklu sınıflandırma URL adresine ait metinsel özellikler ile daha başarılı değerler elde edilmiştir.

Bu çalışmada URL adreslerinin sınıflandırması probleminin çözümüne aşağıdaki katkılar verilmiştir:

- URL adres bilgilerinden elde edilmiş olan özelliklerin kullanıldığı ve gerekli optimizasyon işlemlerinin yapıldığı bir derin sinir ağı tasarlanmıştır.
- Belirli parametrelerin sabit tutularak tekrarlanan deneyler vasıtasıyla modelin kullanılabilirliği ve başarı performansı doğrulanmıştır.
- Metinsel analiz ile elde edilen özelliklerin vektörel özelliklere göre sınıflandırma da daha anlamlı olduğu gösterilmiştir.

Bu çalışmanın 2. bölümünde kötü amaçlı web sayfalarının tespitine yönelik olarak yapılan önceki çalışmalar hakkında bilgi verilmiştir. 3. bölümde, bu çalışmada kullanılan materyaller ve özellik çıkarma ve sınıflandırma metodlarından bahsedilmiştir. 4. bölümde önerilen derin sinir ağının test ortamı, elde edilen sonuçlar karşılaştırmalı olarak gösterilmiştir. Son bölümde ise çalışmanın genel bir özeti yapılarak gelecekte yapılacak çalışmalar hakkında bilgi verilmiştir.

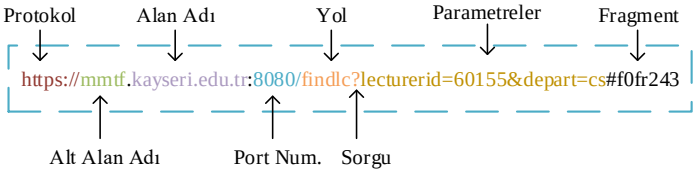
2. Benzer Çalışmalar

Bu bölümde kötü amaçlı web sayfalarının tespitine yönelik olarak çalışmalardan bazılarının kısa özeti verilmiştir. Benzer çalışmalara bakmadan önce URL adreslerinin terminolojisini anlamak önemlidir. Bu sebeple 2.1. bölümünde kısaca URL yapısı anlatılmaktadır. 2.2. bölümünde ise bu çalışmada da kullanılan makine öğrenmesi temelli yaklaşımların kullanıldığı örnek çalışmalar kısaca özetlenmiştir.

2.1. Url Adreslerinin Yapısı

Web adresleri herhangi bir sunucu üzerinde barınmakta olan web sayfalarını tanımlamak için kullanılırlar(Gupta ve ark., 2021). Şekil-1'de örnek bir web adres yapısı gösterilmiştir. Protokol, alt alan adı, alan adı, port numarası, yol, parametreler

ve fragment olmak üzere 7 bölümden oluşmaktadır. Protokol, web tarayıcısının ilgili web sayfasının bulunduğu sunucu ile nasıl iletişim kurması gerektiğini tanımlar. Https, ftp, tcp, udp, ssh en bilinen protokollerden bazılarıdır. Alan adı, bir web sayfasının internet ortamındaki benzersiz tanımı yapar. Yol, bir web sunucusunda ulaşılmak istenen dosyanın veya klasörün bulunduğu konumu (Ör: c:/files/lecturepage.html) göstermektedir. Alt alan adı, bir üst düzey alan adının alt bölümlerinin(Ör. mmtf.kayseri.edu.tr) adıdır. Sorgu bölümü genellikle dinamik çalışan web sayfalarında bulunur. Bir kullanıcı bir sunucuda herhangi bir sayfayı istekte bulunduğunda bir sorgu dizesi alınır ve sunucuda araştırılır. İlgili sayfa bulunur ise kullanıcıya bu sayfa cevap olarak gönderilir.



Şekil 1 Web Adres Yapısı (Gupta ve ark., 2021)

2.2. Kötücül Web Sayfalarının Tespitine Yönelik Olarak Yapılan Güncel Çalışmalar

Son zamanlarda birçok çalışmada kötücül web sitelerinin tespitinde geniş ölçekli veri tabanları ve makine öğrenimi yaklaşımları önerilmiştir. URL, sayfa içeriği, DNS bilgileri gibi çeşitli özellikler çıkarılmakta ve bu özellikleri içeren özellik vektörleri üretilmektedir. Sonrasında makine öğrenmesi algoritmalarının eğitim ve test aşamalarında kullanılmaktadırlar. Sonuçlar seçilen özellik kümesi ve makine öğrenmesi algoritmasına göre değişkenlik göstermektedir.

Li ve ark. (Li ve ark., 2019) , URL ve html özelliklerini kullanarak kimlik avına yönelik web sitelerinin sınıflandırılması için gradyan artırma, karar ağacı, XGBoost ve LightGBM algoritmalarının hibrit olarak kullanıldığı bir model önermişlerdir. Bir tür yığınlama modelidir. Modeli 50 binden fazla web sayfası içeren bir veriseti ile test ederek değerlendirmişlerdir. Modelin %97,30 doğrulukla sınıflandırma yaptığı, %1,16 yanlış pozitif değerinin elde edildiği iddia edilmektedir.

Benzer şekilde URL adresleri ve web içeriği arasındaki kavram bağlantıyı ve tutarlığı doğrulayarak saldırılara karşı koyabilecek model Nureni ve ark. (Nureni ve ark., 2017) tarafından önerilmiştir. PhishDetect isimli yöntem uygulanarak %99,1 doğruluk elde edilirken, naive Bayes algoritması sınıflandırıcı olarak kullanılmıştır.

Şahingöz ve ark. (Şahingöz ve ark., 2019) URL dilinden ve üçüncü taraf bilgilerinden bağımsız olarak çalışan bir model önermişlerdir. Gerçek zamanlı olarak kimlik avı sitelerini tespit edebilmektedirler. Kelime sayısı, marka adı sayısı gibi doğal dil işlemi ile elde edilen toplam 27 özellik sınıflandırma için kullanılmıştır. Rassal orman algoritması ile %97,98 doğruluk elde edildiği iddia edilmiştir.

Evrişimli sinir ağlarının kullanıldığı bir diğer çalışmada ise kötücül URL adreslerinin %100 seviyesinde bir doğrulukla tanımanın mümkün olduğu iddia edilmektedir (Wei ve ark., 2020). Sadece URL metni analiz edilmektedir. Bu sebeple oldukça hızlı bir tespit mekanizması sunar ve sıfırıncı gün saldırıları içinde etkin bir yöntem olduğu belirtilmiştir. Vinayakumar ve ark. (Vinayakumar ve ark., 2018) tekrarlayan sinir ağı, uzun kısa süreli bellek, evrişimli sinir ağı ve uzun kısa süre bellek yapsında evrişimli sinir ağı(hibrit) gibi derin öğrenme mimarilerini kötücül Url tespitinde kullanmıştır. Karakter düzeyinde özellik çıkararak modelleme yapmıştır. Tüm modelleri karşılaştırmalı olarak göstermiştir. Sonuçta LSTM tabanlı modellerin daha başarılı oldunu göstermiştir.

Sonuçta kötücül web sitelerinin tespit edilmesine yönelik olarak yapılan çalışmalarda derin öğrenme modelleri genellikle geleneksel modellerden daha iyi sonuçlar vermektedir. Bu sebeple bu çalışmada önerilen model için derin öğrenme yapısında bir tasarım yapılmıştır ve başarılı sonuçlar elde edilmiştir. Bu sonuçlar elde edilirken herhangi bir görsel semantik veri, içerik analizi yapılmamıştır. Böylece daha az zaman harcayarak yüksek performans elde edilmiştir.

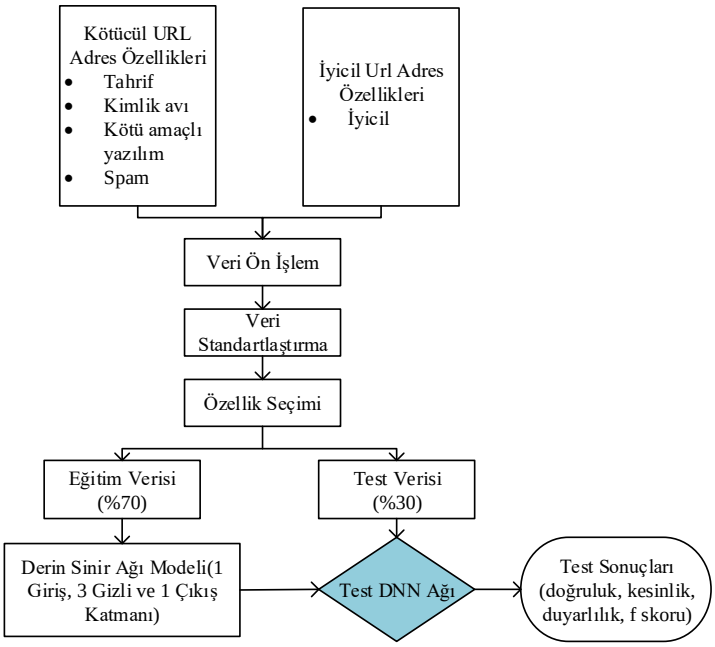
3. Materyal ve Metodlar

3.1. Önerilen Metot

URL filtreleme modelinin akış diyagramı şekil-2'de gösterildiği gibidir. Modelin ilk aşamasında kötücül ve iyicil URL adreslerinin toplanması ve sınıflanlarının belirlenmesi gereklidir. Bu aşamada ISCX veriseti kullanılmıştır. Bu verisetinde 4 kötücül ve 1 iyicil sınıfta web adresleri bulunmaktadır. Veriseti incelendiğinde url adreslerine ait özelliklere ait değerlerde inf, nan gibi işlemesi mümkün olmayan değerlerin bulunduğu görülmüştür. Ayrıca bir takım verilerde de eksik değerlerin olduğu tespit edilmiştir. Özeliklere ait eksik ve işlenemeyen değerler öğrenme aşamasında problem oluşturmaktadır. Bunu aşmak için ilk aşamada bu değerler tespit edilmiş ve verisetinden elenmiştir. Bunun yanında standart programlama dillerinde bulunan veri tipleri ile tutulamayacak kadar büyük değerlere rastlanılmıştır. Bu değerler ile çalışmak için string yapısı kullanılması gerekmiştir. Bunu aşmak için değerlerde standardizasyon işlemi yapılmıştır. Böylece öğrenme aşamasında işlenebilir bir veri kümesinin oluşturulması sağlamıştır.

Veri standartlaştırma sonrasında verisetinde bulunan ve her bir Url için çıkarılmış olan 79 adet özellik incelenmiş olup bu özellikler üzerinde özellik seçim işlemleri yapılmıştır. Böylece model üzerinde etkisiz ve bozucu etkisi olacak özelliklerin elenmesi, hızlı bir eğitim ile daha yüksek başarı düzeyi yakalanması amaçlanmıştır.

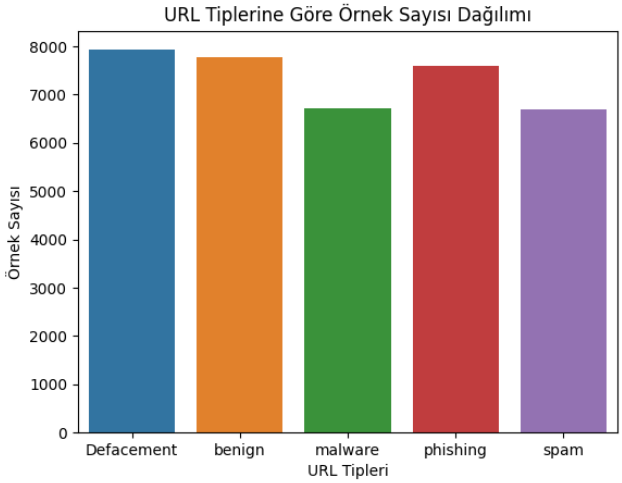
Özellik seçimi sonrasında eğitim ve test aşamasında kullanmak üzere veriler her bir sınıf için ayrı ayrı olacak şekilde %70-%30 oranında ayrılmış ve iki ayrı özellik vektör yapısı oluşturulmuştur. Eğitim kümesi ile derin sinir ağının eğitimi gerçekleştirildikten sonra test verisi ile bu ağın sınıflandırma performansı ölçülmüştür. Ölçüm sonuçları farklı metriklere göre değerlendirilmiştir. Ayrıca çok sınıflı bu problem için hangi sınıflarda sorun yaşandığını anlamak için testlerde karmaşıklık matrisleri de üretilmiş ve karşılaştırması yapılmıştır.



Şekil 2 Önerilen Modelin Akış Diyagramı

3.2. Veriseti Detayları

Bu çalışmada önerilen modelin test edilmesi ve değerlendirilmesi için Canada Brunswick üniversitesi tarafından dağıtımı yapılan ISCXURL-2016 veriseti (Mamun ve ark., 2016; ISCX-URL Dataset, 2016) kullanılmıştır. Verisetinin örnek dağılımı Şekil-2'de gösterilmiştir. Tahrif saldırısı kategorisinde 7930, iyicil türünde 7781, kimlik avı saldırısı türünde 7586, kötü amaçlı yazılım saldırısı türünde 6712 ve spam türünde 6698 örnek bulunmaktadır. Kümeler arasında örnek sayılarında kısmen dengesizlik olmak ile birlikte sınıflandırma sonuçlarına etki edecek düzeyde değildir. Bu sebeple sonuçlarda objektifliği sağlamak adına veri ile ilgili olarak herhangi bir veri artırma (data augmentation) algoritması uygulanmamıştır



Şekil 3 ISCX URL 2016 Veriseti Örnek Dağılımı

Tahrif(defacement) saldırısı, kötü niyetli saldırganların bir web sitesine girip sitedeki içeriği kendi mesajları ile değiştirdiği atağı ifade etmektedir. Mesajlar web sitesi sahiplerini utandıracak siyasi ve dini mesajlar, küfür veya uygunsuz içerikler olabilmektedir. Web sitesinin görünen içeriğini etkileyen, şablon yada sayfa içeriğinin bulunduğu dosyalarda meydana gelen beklenmeyen değişiklikler tahrif saldırısına işaret etmektedir (Mao ve ark., 2019).

Kimlik avı saldırısı(phishing) yapan Url türü, hassas kişisel bilgilere erişmek için kimlik avı gerçekleştirir. Kullanıcıların oturum açma kimlik bilgileri, hesap numaraları, pin numaraları, kredi kartı bilgileri gibi çeşitli sosyal mühendislik teknikleri ile elde etmeyi amaçlayan web içeriğe sahiptir. Bir Url adresine tıklanması, bir epostanın açılması ile ortaya çıkabilmektedir (Chiramdasu ve ark., 2021).

Kötü amaçlı yazılım saldırısı(malware attack) son derece kötü amaçlı dosyalar yada programlardır. Bilgisayar virüseri, truva atları, fidye yazılımları, casus yazılımlar bu gruba girmektedir. Hedef sisteme yetkisiz erişim, ele geçirme, kullanıcı etkinliklerini izleme gibi kötücül faaliyetleri gerçekleştirirler (Chiramdasu ve ark., 2021).

Spam türü ise kullanıcılara çok sayıda mesaj göndermek için bir uygulamanın yetkisiz olarak kullanılmasıdır. Bu mesajlar sahte veya saldırıya uğramış profiller tarafından gönderilmektedir. Genellikle gerçek kullanıcıların gerçek dışı reklamları ve bağlantıları tıklanması istenilmektedir. Verisetinde farklı türlerde kötücül Url adresleri bulunması önerilen modelin testlerinde objektifliği yakalamak için önemlidir. Ancak her bir kötücül yazılım türünde alınacak önemler birbirinden farklı olacağı için kötücül yazılım türlerinin belirlenmesi faydalı olmaktadır (Manyumwa ve ark., 2020).

Verisetinde bulunan her bir Url adresi için 79 farklı özellik çıkarılmıştır. Bu özelliklerde bazıları; sorgu uzunluğu, dosya uzantısı, url karakter sayısı, domain karakter sayısı, yol karakter sayısı gibidir. Özelliklere ilişkin detaylar Mamun ve ark. tarafından yapılan çalışmada gösterilmiştir(Mamun ve ark., 2016). Bu çalışmada her bir Url adresine ait 79 özellik içerisinden özellik seçim metodları kullanarak sadece anlamlı özellik seçilmiş olup derin sinir ağı eğitiminde kullanılmıştır

3.3. Değerlendirme Metrikleri

Bu çalışmada 4 tür kötücül ve 1 tür iyicil URL adreslerinin sınıflandırması yapılmıştır. En iyi performans gösteren modeli doğrulamak için karışıklık matrisini; modelin performansını değerlendirmek için kesinlik, duyarlılık, doğruluk ve f skoru metrikleri kullanılmıştır.

- Karmaşıklık matrisi: Bir sınıflandırıcının doğruluğunu değerlendirmek için kullanılmaktadır. Tahmin edilen ve gerçek sınıflandırma değerlerini içerir.

$$\text{Karmaşıklık matrisi} = \begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} \quad (1)$$

TP, doğru olarak tahmin edilen kötücül URL sayısını; TN, doğru olarak tahmin edilen iyicil URL sayısını; FP yanlış olarak tahmin edilen kötücül URL sayısını ve FN, yanlış olarak tahmin edilen iyicil URL sayısını göstermektedir.

- Doğruluk: doğru olarak tahmin edilen kötücül URL sayısının, toplam örnek sayısını oranıdır. Hesaplama formülü aşağıdaki gibidir (Eşitlik-2):

$$\text{Doğruluk} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (2)$$

- Kesinlik: doğru olarak tahmin edilen kötücül URL sayısının, tüm kötücül URL sayısına oranını ifade eder. Hesaplama formülü aşağıdaki gibidir (Eşitlik-3):

$$\text{Kesinlik} = \frac{(\text{TP})}{(\text{TP} + \text{FP})} \quad (3)$$

- Duyarlılık: doğru olarak tahmin edilen kötücül URL sayısının, yanlış olarak tahmin edilen iyicil URL sayısına oranıdır. Hesaplama formülü aşağıdaki gibidir (Eşitlik-4):

$$\text{Duyarlılık} = \frac{(\text{TP})}{(\text{TP} + \text{FN})} \quad (4)$$

- F Skoru: Kesinlik ve recall değerlerini birlikte değerlendirmek için kullanılan değerdir. İki değerin harmonik ortalamasıdır. Hesaplama formülü aşağıdaki gibidir (Eşitlik-5):

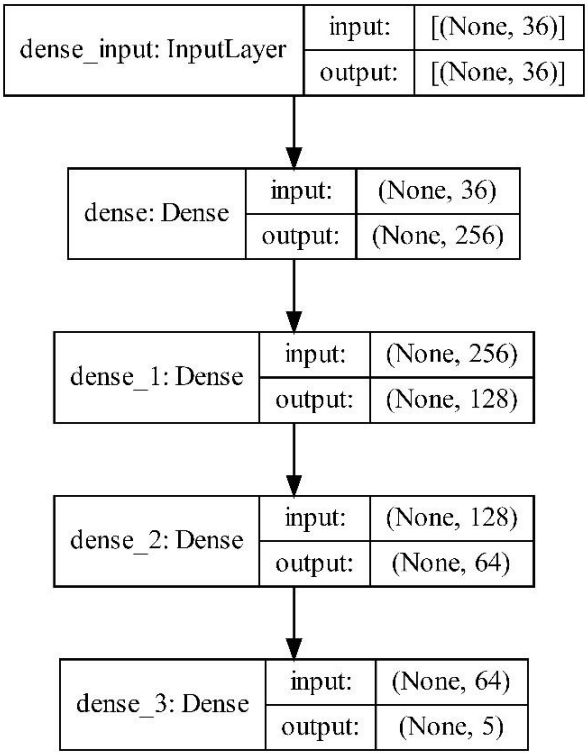
$$\text{F skoru} = \frac{(2 \text{ x Kesinlik x Duyarlılık})}{(\text{Kesinlik} + \text{Duyarlılık})} \quad (5)$$

Yüksek doğruluk değeri modelin genel olarak tüm URL tipleri için başarılı bir sınıflandırma yeteneğine sahip olduğunu gösterir. Yüksek duyarlılık değeri yüksek olursa, ilgili tür için o kadar başarılı tanıma yaptığını ifade eder. Kesinlik değeri yüksek olursa, ilgili tür için o kadar yüksek doğruluk değerinin elde edildiğini garanti eder. F skoru değeri yüksek olursa da sistemini tüm türler için sınıflandırma sonuçlarındaki kararlılığının o kadar yüksek olduğunu gösterir.

3. Deneysel Sonuçlar ve Tartışma

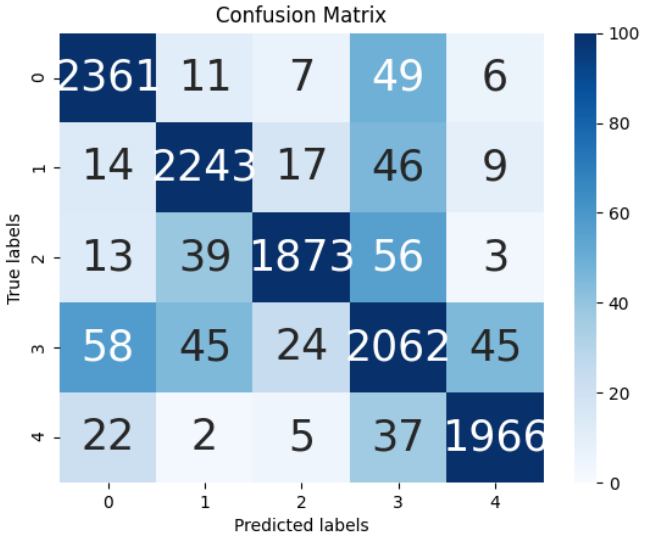
Bu çalışmada önerilen model testlerinde Tensorflow'u Keras kütüphanesi ile birlikte kullanılmıştır. Derin öğrenme mimarisinin GD (Gradient Descent) hızının hesaplama hızlını artırmak için GPU tabanlı bir ekran kartı kullanıldı. Deneylerdeki parametrelerin sürekli olarak ayarlanması ve optimize edilmesi için etkin bir model tasarımı yapıldı. Şekil'4'te gösterildiği gibi giriş katmanında 36 özelliğin girdi olarak kullanılmıştır. 3 gizli katmanda sırasıyla 256, 128 ve 64 düğüm bulunmakta olup çıkış katmanında 1 iyicil ve 4 kötücül olmak üzere 5 sınıf bulunmaktadır. Optimizasyon algoritması olarak adam algoritması kullanılmıştır. Gizli katmanlar arasında herhangi bir seyreltme işlemi yapılmamıştır. Önerilen yaklaşımın uygulanmasında Python dili kullanılmıştır. Scikit-learng, Scipy, Pandas, Matplotlib kütüphanelerinden yararlanılmıştır.

Bu çalışmada önerilen derin sinir ağı modeli 5 sınıftan oluşan Url filtreleme probleminde ortalama olarak %95,4 doğruluk ile sınıflandırma yapabilmektedir. Birinci çalışmamıza göre ortalama olarak %7,3'lük bir gelişme sağladık. Modelin derin katman sayısı artırıldığında eğitim süresi uzamakta ancak sınıflandırma sonucunda herhangi değişiklik olmamıştır. En optimum modele ulaşabilmek için farklı katman sayısı ve katmanlardaki farklı düğüm sayıları ile testler gerçekleştirilmiştir. Sonuçta Şekil-4'te verilen model ile en iyi değerler elde edilmiştir



Şekil 4 Önerilen Model Derin Sinir Ağı Model Yapısı

Modelin sınıf bazında sonuçlarını gözlemleyebilmek için karmaşıklık matrisi Şekil-5'te gösterilmiştir. Görüleceği üzere doğru tahmin edilen örnekler çoğunlukla köşegen üzerinde yoğunlaşmaktadır. Bu da %95,4 doğruluk değerini kanıtlamaktadır. Diğer taraftan sınıflar arasında örnek sayısında dengesizlik olmasına rağmen tüm sınıflarda başarı yakalanmıştır. Bu da modelin Url tiplerinden bağımsız olarak kötücül türleri kendi arasında sınıflandırabildiğini göstermektedir. Ayrıca modelin kimlik avı sınıfını, diğer sınıflar ile karıştırma konusunda eğilimi vardır. Verisetinde bulunan örnek sayısının artırılması halinde model öğrenme düzeyinin artacağı ve sınıflar arası ayrımın daha net şekilde yapılacağı düşünülmektedir.



Şekil 5 Url Sınıflandırma Modeli Karmaşıklık Matrisi

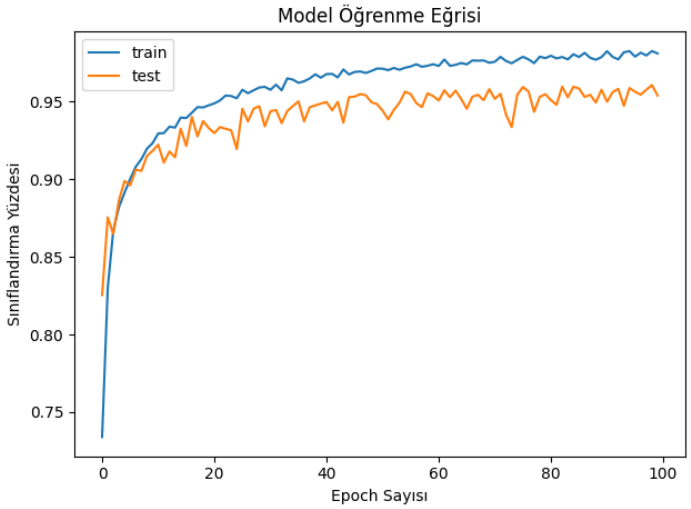
Tablo-6'da gösterilen performance matrisine bakıldığında 5 farklı Url tipinin kesinlik değeri %92 ile %97 arasında değişen düzeylerde olmuştur. Benzer şekilde duyarlılık değeri de %92 ile %97 arasında sınıflara göre değişkenlik göstermektedir. F skoru

değeri ise kötü amaçlı yazılım türü haricinde %97 olmuştur. Sonuçta makro ortalama değeri %95,4 düzeyinde olmuştur. Bu noktada kötü amaçlı yazılım türü(malware) sınıfında başarı düzeyi diğer sınıflara göre düşük kalmıştır. Bu da genel performansı olumsuz etkilemektedir. Bunun sebebinin ilgili sınıfta örnek sayısının az olması olduğu düşünülmektedir. Bu konuda daha detaylı çalışma yapılacak olup modelin genel performansını etkileyen hususlar ayrıca değerlendirilecektir.

Tablo 1 URL Filtreleme Modeli Performans Matrisi

	Kesinlik	**Duyarlılık**	**F skoru**	**Destek**
Tahrif saldırısı	0,96	0,97	0,96	2434
İyicil	0,96	0,96	0,96	2329
Kimlik avı	0,97	0,94	0,96	2234
Kötü amaçlı yazılımlar	0,92	0,92	0,92	1984
Spam	0,97	0,97	0,97	2032
Doğruluk			0,95	11013
Makro ortalama	0,95	0,95	0,95	11013
Ağırlıklandırılmış ortalama	0,95	0,95	0,95	11013

Son olarak bu çalışmada önerilen derin sinir ağı modelinin öğrenme eğrisi de Şekil-7'de gösterilmiştir. Buna göre model 100 Epoch'tan itibaren öğrenme sürecinin tamamlamaktadır. Eğitim ve test modelleri paralel bir şekilde ilerlemekte olup modelin kararlılığını doğrulamaktadır.



Şekil 6 Model Öğrenme Eğrisi

4. Sonuçlar

Bu çalışma kötücül URL adreslerinin tespitine yönelik olarak çoklu sınıflandırma yapan bir derin sinir ağı modelini tanıtmaktadır. Öncelikle mevcut yaklaşımlara ilişkin bilgiler verilmiştir. Sonrasında Url filtreleme problem genel hatları ile tanımlanmıştır. ISCX URL 2016 veristinin kullanıldığı bu devam çalışmasında birinci çalışmamızda çoklu sınıflandırma için çözüm bulamadığımız duruma metin tabanlı analiz yaparak çözüm bulmaya çalıştık. Yaklaşımımızda URL adresine ait 36 metinsel özellik kullanıldı ve sonuçta ortalama olarak %95,4 sınıflandırma başarısı yakalandı. Önerilmiş olan DNN modeli bu konudaki bir önceki çalışmamızla birlikte farklı metriklerle değerlendirildi ve sonuçları karşılaştırmalı olarak gösterildi.

Gelecekte yaklaşım, standart DNN modeli yerine daha fazla geri besleme imkânı tanıyan LSTM yapıları ile değerlendireceğiz. Ayrıca yüksek başarı düzeyini korurken, bellek ve işlem zamanında gerekli azalmayı sağlamak için optimizasyon araştırması yapacağız.

Kaynakça

Arslan, R.S. (2021). Kötücül web sayfalarının tespitinde Doc2Vec modeli ve makine öğrenmesi yaklaşımı. European Journal of Science and Technology (Accepted).

Chiramdasu, R., Srivastava, G, Bhattacharya, S., Reddy, P.K. & Gadekallu, T.R. (2021, Ağustos). Malicious URL Detection using Logistic Regression. International Conference on Omni-Layer Intelligent Systems (COINS)(pp. 467-482).

Devi, M. I., Selvakuberan, K. & Rajaram, R. (2009). Fast web page classification without accessing the web page using machinelearning techiques. Journal of Information, Intelligence and Knowledge, 1(1), 1-10.

Ebeling, J.(2021, 17 Şubat). Domain Age as an Internet Filter Criteria. Erişim adresi https://www.mcafee.com/blogs/enterprise/cloud-security/domain-age-as-an-internet-filter-criteria/

Gupta, B.B., Yadav, K., Razzak, I., Konstantinos, P., Castiglione A. & Chang, X. (2021). A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment. Computer Commmunications, 175, 47-57.

He, S., Li, B., Peng, H., Xin, J. & Zhang, E. (2021). An effective cost-sensitive Xgboot method for malicious URLs detection in imbalanced dataset. IEEE Access, 9, 1-8.

Hou, Y., Chang, Y., Chen, T. & Laih, C. (2010). Malicious web content detection by machine learning. Expert systems with applications 37(1), 55-60.

Huang, Y., Li, T., Zhang, L., Li, B. & Liu, X. (2020). JSContana: Malicious Javascript detection using adaptable context analysis and key feature extraction. Computer & Security, 104, 1-9.

ISCX-URL2016 legitimate and phishing URL Dataset (2021, 1 Eylül). Erişim adresi https://www.unb.ca/cic/datasets/url-2016.html

Li, Y., Yang, Z., Chen, X., Yuan, H. & Wenyin, L. (2019). A stacking model using URL and Html features for phishing web page detection. Future Generation Computer Systems, 94, 27-39.

Malware variety grows by 13.7% in 2019 due to web skimmers. (2019, 10 Aralık). Erişim adresi https://www.kaspersky.com/about/press-releases/2019_malware-variety-grows-by-137-in-2019-due-to-web-skimmers

Mamun, M. S. I., Rathore, M. A., Nowak, J., Lashkari, A.H., Stakhanova, N. & Ghorbani, A.A. (2016, Eylül). Detecting Malicious URLs Using Lexical Analysis. International

Conference on Network and System Security (pp. 467-482).

Mao, B.M., & Bagolibe, K.D. (2019, Ekim). A Contribution to Detect and Prevent a Website Defacement. 2019 International Conference on Cyberworlds (CW) (pp. 1-4).

Manyumwa, T., Chapita, P.F., Hanlu W. & Ji, S. (2020, Aralık). Towards Fighting Cybercrime: Malicious URL Attack Type Detection using Multiclass Classification. International Conference on Big Data (Big Data) (pp. 1813-1822).IEEE.

Okunoye, O.B., Nureni, A.A. & Illurimi F.A (2017). A Web Enabled Anti-Phishing Solution using Enhance Heuristic Based Technique. Journal of Research in Sciences, 13(2), 304-321.

Palaniappan, G., Sangeetha, S., Rajendran, B., Goyal, S.S. & Bindhumadhava, B. S. (2020, Aralık). Malicious domain detection using machine learning on domain name features, host-based features and web-based features. In 2019 3rd international conference on computing and network communications (CoCoNet'19) (654-661).

Ravi, V., Soman, K.P. & Pornachandran, P. (2018). Evaluationg deep learning approaches to characterize and classify malicious URL's. Journal of Intelligent and Fuzzy Systems, 34, 1333-1343.

Stevanoviz, M., Pedersen, J. M., D'Aconzo, A., Ruehrup, S. & Berger, A. (2015). On the ground truth problem of malicious DNS traffic analyis. Computer & Security, 55, 142-158.

Şahingöz, Ö. K., Buber, E., Demir, Ö. & Diri, B. (2019). Machine learning based phishing detection from URLs. Expert Systems with Applications, 117, 345-357.

Wei, W., Qiao, K., Nowak, J., Korytkowski, M., Scherer, R. &Wozniak, M. (2020). Accurate and fast URL phishing detector: A convolutional neural network approach. Computer Networks, 178, 1-9.

Yuan, J., Chen G., Tian, S. & Xinjun, P. (2021). Malicious URL detecion based on a parallel neural joint model. IEEE Access, 9, 1-9.