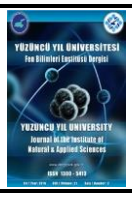
Yüzüncü Yıl Üniversitesi
Fen Bilimleri Enstitüsü Dergisi

https://dergipark.org.tr/tr/pub/yyufbed



Araştırma Makalesi

Öğrencilerin Siber Güvenlik Farkındalık Düzeylerinin Makine Öğrenmesi Yöntemleri ile Belirlenmesi

Mahmut TOKMAK

Mehmet Akif Ersoy Üniversitesi, Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksekokulu Fakültesi, Yönetim Bilişim Sistemleri Bölümü, 15300, Burdur, Türkiye
Mahmut TOKMAK, ORCID No: 0000-0003-0632-4308
Sorumlu yazar e-posta: mahmuttokmak@mehmetakif.edu.tr

Makale Bilgileri

Geliş: 29.09.2022
Kabul: 16.03.2023
Online Ağustos 2023

DOI:10.53433/yyufbed.1181694

Anahtar Kelimeler
Makine öğrenmesi,
Siber güvenlik,
Siber güvenlik farkındalık düzeyi

Öz: Bilgi ve iletişim teknolojilerinin hızla gelişmesi ile birlikte teknoloji ve interneti kullanan cihaz sayısı artmış ve hayatın her alanına girmiştir. Teknolojideki gelişmeler kullanıcıların ve cihazların siber tehditlerle karşılaşma riskini de beraberinde getirmiştir. Bu çalışma; siber tehditlerle ilgili, öğrencilerin siber güvenlik farkındalık düzeylerini makine öğrenme yöntemleri ile tespit etmeyi amaçlamaktadır. Bu nedenle istatistiksel olarak lisans öğrencilerini temsil eden örnek bir kitleden anket tekniğiyle veri toplanmıştır. Elde edilen veriler, betimsel tarama modeli benimsenerek analiz edilmiş ve analiz sonuçları çalışmada ortaya konmuştur. Sonrasında anket verilerinden oluşturulan veri seti ile Naive Bayes, Karar Ağacı, Rastgele Orman, En Yakın Komşu, XGBoost, Gradient Boost, Destek Vektör Makineleri, Çok Katmanlı Algılayıcı algoritmaları kullanılarak öğrencilerin siber güvenlik farkındalık düzeylerinin tespiti yapılmıştır. Yapılan testler sonucunda 0.7-0.98 arasında değişen doğruluk değerleri, 0.7-0.96 arasında değişen F1 skorları elde edilmiştir. En başarılı performans metrikleri 0.98 doğruluk ve 0.96 F1-skoru ile Çok Katmanlı Algılayıcı algoritması ile elde edilmiştir.

Determination of Cyber Security Awareness Levels of Students with Machine Learning Methods

Article Info

Recieved: 29.09.2022
Accepted: 16.03.2023
Online August 2023

DOI:10.53433/yyufbed.1181694

Keywords
Cyber security,
Cyber security awareness level,
Machine learning

Abstract: With the rapid development of information and communication technologies, the number of devices using technology and the internet has increased and has entered all areas of life. Developments in technology have brought the risk of users and devices encountering cyber threats. This work aims to determine students' cyber security awareness levels about cyber threats with machine learning methods. Therefore, data were collected from a sample population that was statistically representative of undergraduate students with the survey technique. The obtained data were analyzed by adopting the descriptive review model and the results of the analysis were presented in the study. Afterwards, the cyber security awareness levels of the students were determined by using the data set created from the survey data, Naive Bayes, Decision Tree, Random Forest, Nearest Neighbor, XGBoost, Gradient Boost, Support Vector Machines, Multi-Layer Perceptron algorithms. As a result of the tests performed, accuracy values ranging from 0.7-0.98 and F1 scores ranging from 0.7-0.96 has been obtained. The most successful performance metrics were obtained with the Multi-Layer Perceptron algorithm with an accuracy of 0.98 and an F1 score of 0.96.

1. Giriş

Bilgi Teknolojisi (BT) hayatımızın her alanına girmiştir. BT bize insan yaşamının her yönünü zenginleştiren fırsatlar ve kolaylıklar sağlamaktadır. Bugün, BT'nin olmadığı bir dünya pek hayal edilemez durumdadır. Öyle ki 2022 yılında dünya nüfusunun %69'u ve Türkiye nüfusunun %83.8'i internet kullanıcısı olarak istatistiklere geçmiştir (IWS, 2022). Türkiye İstatistik Kurumu'nun 2021'de yayınladığı raporda da 16-74 yaş aralığındaki bireylerin internet kullanma oranının %3.6 oranında arttığı bildirilmiştir. Ayrıca aynı raporda E-devlet kullanımının %58.9 olarak gerçekleştiği ve bir önceki yıla göre %7.4 oranında artış gösterdiği, internet üzerinden mal veya hizmet satın alma oranının %44.3 olduğu ve bir önceki yıla göre %7.8 oranında arttığı bildirilmiştir (TÜİK, 2021). Bunun yanı sıra mobil cihazlardaki artış ve nesnelerin interneti teknolojisiyle birlikte kullanıcılar normal dünya yaşamının yanı sıra siber dünyaya da entegre olmaya başlamışlardır (Karacı ve ark., 2017). Ancak, fırsatlar yanında zorluklar da getirebilmektedir. Siber dünyadaki bu gelişmeler, siber saldırganların ve siber saldırıların ortaya çıkması ve günden güne sayısının artması gerçeğini ortaya çıkarmıştır.

Siber güvenlik bilinci günümüzde her zamankinden daha önemli bir hale gelmiştir. Kişisel bilgilere yönelik tehditlerin arttığı ve kişisel bilgilerin her gün çalındığı, zararlı yazılımların kişilere ve kurumlara verdiği zararlar görülmektedir. Bireylerin bu konuda bilinçlendirilmesi ilk adım olarak kabul edilmektedir. Siber güvenlik farkındalık düzeyini tespit edip, uygun farkındalık programını benimseyerek saldırıların etkisi azaltılabilir (Subramaniam, 2017).

Siber güvenliğin sağlanması hususunda Makine Öğrenmesi (Machine Learning: ML) gibi teknolojik yöntemlerle önlemler alınabilmektedir. Ancak siber güvenlik alanında bir diğer faktörün de insan faktörü olduğu göz ardı edilmemelidir. Birçok siber saldırıya maruz kalmanın ve siber güvenlik zafiyetlerinin sebebi olarak insan hataları gösterilmektedir. Dolayısıyla insan faktöründeki hataları en aza indirmek için siber güvenlik farkındalığının kazandırılması gerekmektedir (Yiğit & Seferoğlu, 2019). Siber güvenlik farkındalığı "bilgi güvenliği kavramının önemi, bir kuruluşa ait verileri ve ağları korumak amacıyla yeterli düzeyde bilgi kontrolü ve uygulama sorumluluklarını anlama derecesi" olarak tanımlamıştır. Siber güvenlik farkındalığı kavramının amaçları arasında; interneti kullanan kişileri siber güvenlik riskleri hakkında uyarmak ve kullanıcıları internet kullanımı sırasında güvenliği benimsemeye yeterince kararlı olmaları için siber güvenlik riskleri konusundaki anlayışlarını geliştirmek vardır. Bu nedenle, siber güvenlik farkındalığı faktörü insan kaynaklı hataların veya güvenlik açıklarının azaltılması, güvenliğin kişisel veya kurumsal düzeyde iyileştirilmesinde kilit bir faktördür (Quayyum ve ark., 2021).

Bu çalışmada; Mehmet Akif Ersoy Üniversitesi, Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksekokulu bölümlerinde öğrenim gören öğrencilerin siber güvenlik farkındalık düzeylerinin ML yöntemleri ile tespit edilmesi hedeflenmiştir. Bu amaçla Siber Güvenlik Ölçeği (SGÖ) kullanılarak, anket yoluyla veriler toplanmış ve bu veriler kullanılarak bir veri seti oluşturulmuştur. Bu nedenle çalışmada anket yoluyla elde edilen veriler için öncelikle, öğrencilerin siber güvenlik farkındalık düzeyleri ile ilgili mevcut durumun betimlenmesi yapılmış ve değişkenler arası ilişkiler incelenmiştir. Bunun yanı sıra siber güvenlik farkındalık düzeylerinin bazı değişkenlere göre farklılık gösterme durumu ortaya konmuştur. İstatitistiksel incelemelerde betimsel tarama modeli benimsenmiştir (Fraenkel ve ark., 2012). İstatistiki olarak bulguların ortaya konmasından sonra Naive Bayes (NB), Karar Ağacı (Decision Tree: DT), Rastgele Orman (Random Forest: RF), K-En Yakın Komşu (K-Nearest Neighbors: KNN), XGBoost, Gradient Boost, Destek Vektör Makineleri (Support Vector Machine: SVM), Çok Katmanlı Algılayıcı (Multi-Layer Perceptron: MLP) ML teknikleri ile siber güvenlik farkındalık düzeylerinin sınıflandırılması yapılmıştır. Kurulan ML modelleri eğitilip teste tabi tutulduktan sonra performans metrikleri ortaya konmuştur. Çalışma, öğrencilerin siber güvenlik farkındalık düzeylerinin ML teknikleri ile kolay ve hızlı bir şekilde tespit edilmesi açısından önemlidir. Siber güvenlik farkındalık düzeylerinin ortaya konması adına istatistiksel yöntemlerle yapılan çalışmalara ek olarak, ML teknikleri ile bu farkındalık düzeylerinin belirlenmesi, yapılan eğitim ve testler sonucu elde edilen bulguların literatüre katkı yapması hedeflenmiştir.

2. Teorik Altyapı

2.1. Siber güvenlik

Siber güvenlik; "siber ortamı, kuruluşu ve kullanıcıya ait varlıkları koruma amacıyla kullanılabilecek olan araçlar, güvenlik amaçlı önlemleri, uygulanan politikalar, yönergeler, risklerin yönetilmesi yaklaşımları, eylemler, verilen eğitimler, geliştirilen uygulamalar ve bu alandaki teknolojilerin toplamı olarak tanımlanmaktadır. Kuruluşlar ile kullanıcıların varlıkları kavramı ise bilgi işlem cihazlarını, çalışan personeli, oluşturulmuş altyapıyı, verilen hizmetleri, geliştirilen uygulamaları, telekomünikasyon sistemlerini ve siber ortamda iletilen ve/veya depolanan bilgi ve belgelerin tamamını kapsamaktadır" (Von Solms & Van Niekerk, 2013; Alzahrani, 2021). Siber güvenlik hedefleri; erişebilirlik", bütünlük ve gizlilik kavramları ile ifade edilmektedir. Erişilebilirlik; bilgi ve bilgi sistemlerinin yetkili kişilerce ulaşılabilir olması olarak tanımlanmaktadır. Bütünlük; bilgilerin yetkisiz düzenlemeye veya yok edilmeye karşı korunması olarak ifade edilmektedir. Buradaki amaç bilgi ve bilgi sistemlerinin doğru, eksiksiz ve bozulmaya uğramamış olmasını temin etmektir. Gizlilik; bilginin yetkisiz olan erişimlere karşı korunması olarak tanımlanmaktadır. Bilgiye erişme hakkına sahip olanların işlem yapabilmelerini, yetkilendirilmemiş kişilerin bu işlemleri yapmalarının engellenmesi olarak tanımlanabilmektedir (Von Solms & Van Niekerk, 2013).

2.2. Siber güvenlik terminolojisi

Siber uzay, en yeni bilgi ve iletişim teknolojilerini kullanarak, birbirine bağlı ve bağımlı ağların yardımıyla bilgi oluşturmak, güncellemek, depolamak, paylaşmak ve kullanmak için elektronik ve elektromanyetik spektrumun kullanımı olan bilgi dünyası içindeki küresel bir alandır (Humayun ve ark., 2020). Bu küresel alanda suç içeren durumlarla ve tehditlerle karşılaşabilmektedirler. Siber güvenlik alanıyla ilgili anahtar kavramların daha iyi anlaşılması için önemli terminolojilerin bazı tanımları, siber suç ve tehditlerin en yaygın bilinenleri aşağıda verilmiştir.

Zafiyetler (Vulnerabilities), bir saldırganın kötü niyetli komutlar yürütmesine, verilere yetkisiz bir şekilde erişmesine ve/veya çeşitli hizmet reddi saldırıları gerçekleştirmesine izin veren, bir sistem veya tasarımındaki hatalardır (Yalçınkaya & Küçüksille, 2021). *Tehditler (Threats),* bir sistemdeki güvenlik açıklarından fayda sağlamak ve sistemi olumsuz yönde etkilemek için yapılan işlemlerdir (Abomhara & Køien, 2015). *Saldırılar (Attacks),* çeşitli araçlar ve teknikler kullanılarak güvenlik açıklarından yararlanmak suretiyle bir sisteme zarar vermek veya rutin işleyişini bozmak için yapılan eylemlerdir. *Saldırganlar (Attackers),* saldırı eylemini gerçekleştiren kişilerdir. Kötü amaçlarına ulaşmak için yaptıkları bu eylemleri kendini tatmin etmek için yada finansal bir kazanç veya ödül için başlatabilmektedirler (Abomhara & Køien, 2015; Humayun ve ark., 2020).

- *Kötü Amaçlı Yazılım (Malware),* "malicious software"'in kısaltmasıdır. Saldırganlar tarafından veri çalmak, bilgisayarlara ve bilgisayar sistemlerine zarar vermek veya onları yok etmek için geliştirilen zararlı yazılımları ifade eder. Bunlardan bazıları: virüs (virus), solucan (worm), casus yazılımlar (spyware), reklam yazılımları (adware), Truva atı (Trojan), botnet, kök kullanıcı takımı (rootkit), arka kapılar (backdoor) olarak adlandırılmaktadırlar (Khan ve ark., 2020; Khan ve ark., 2021).
- *Hizmet Reddi (DoS/DDos)* saldırıları: bir makineyi veya ağ kaynağını hedeflenen kullanıcılar için erişilemez hale getirmeyi amaçlamaktadır. Başlangıçta tek bir kaynaktan saldırı yapılmaktayken şimdilerde aynı anda birden fazla kaynaktan hedef sistemlere saldırılar gerçekleştirilmektedir (İlker, 2019).
- *Oltalama (Phishing)* bir internet kullanıcısından hassas bilgiler toplamak için sosyal mühendislik ve teknoloji kullanan bir saldırı çeşididir. Oltalama teknikleri, e-posta, anlık mesajlar, açılır mesajlar veya Web sayfaları gibi çeşitli iletişim yöntemlerini kullanır (Khonji ve ark., 2013).
- *SQL enjeksiyon (SQL injection)* saldırısında, SQL ifadesini saldırganın lehine değiştirmek veya manipüle için uygulama aracılığıyla bir girdi dizesi enjekte edilir. Bu saldırı, veri

tabanına yetkisiz erişim ve veri tabanının manipülasyonu ve hassas verilerin ifşası dahil olmak üzere çeşitli şekillerde zarar verir. Bu saldırı, yetkisiz gruplar tarafından veri kaybına veya verilerin kötüye kullanılmasına neden olabileceğinden risklidir ve sonuç olarak işlevsellik ve gizlilik yok edilir. Ayrıca, sistem düzeyinde komutlar da bu saldırı kategorisi altında yürütülür ve yetkili kullanıcıların gerekli bilgilere erişememesine neden olur (Humayun ve ark., 2020).

- *Oturum ele geçirme (Session hijacking) ve ortadaki adam (Man-in-the-Middle)* saldırıları Man-in-the-middle (MITM, literatürde MIM, MitM, MiM veya MITMA olarak da kısaltılır), yetkisiz bir üçüncü tarafın birden çok uç nokta arasındaki iletişim kanalının kontrolünü gizlice ele geçirdiği bir saldırıdır. MITM saldırganı, kurbanların iletişim trafiğini kesintiye uğratabilir, manipüle edebilir ve hatta değiştirebilir. Ayrıca, kurbanlar davetsiz misafirin farkında değildir, bu nedenle iletişim kanalının güvenli ve korumalı olduğuna inanırlar.
- *Siteler Arası Komut Dosyası (XSS)* Bu tür saldırıda, kötü niyetli bir saldırgan, müşterinin hassas verilerini çalmak için istemcinin tarayıcısında bir JavaScript kodu çalıştırmaya çalışır (Weamie, 2022).
- *Sıfır gün saldırısı (Zero-day attack)*, yamanın yayınlanmadığı veya uygulama geliştiricilerin habersiz olduğu, bilinmeyen bir güvenlik açığı tehdidini tanımlamak için kullanılan terim olarak kabul edilir (Sarker ve ark., 2020).

2.3. Makine öğrenme yöntemleri

Yapay zekanın alt alanı olarak tanımlanan makine öğrenimi tekniklerinin uygulamaları, eğitim, tıp, finans, imalat endüstrisi, otomotiv gibi yaşamın farklı alanlarında kullanılmaktadır. Bu alanlardan biride siber güvenlik alanıdır. Makine öğrenimi teknikleri, ağın her iki tarafında, yani saldıran tarafta ve savunan tarafta rol oynamaktadır. ML teknikleri saldıran tarafta, savunma duvarını geçmek için kullanılır. Buna karşılık, savunma tarafında, ML teknikleri hızlı ve sağlam savunma stratejileri oluşturmak için uygulanmaktadır. Makine öğrenimi teknikleri, izinsiz giriş tespit sistemi, kötü amaçlı yazılım tespiti, kimlik avı tespiti, spam tespiti ve dolandırıcılık tespiti gibi siber güvenlik tehditlerine ve saldırılara karşı mücadelede hayati bir rol oynamaktadır (Shaukat ve ark., 2020). Çalışmada kullanılan ML yöntemleri teorik olarak aşağıda verilmiştir.

- *Naive Bayes* sınıflandırıcısı, Bayes teoremine dayanan olasılıksal bir yaklaşımdır. Tüm nitelikler veya parametreler istatistiksel olarak birbirinden bağımsız olmalıdır. Mevcut etiketlenmiş örnek verileri kullanarak, sınıfı belirlenecek yeni verilerin hangi sınıfa dahil edileceğinin olasılığını hesaplar (Berrar, 2018).
- *Karar Ağacı,* normal bir ağaçta olduğu gibi karar ağaçları da kök düğümü, dallar ve yaprak düğümlerinden oluşur. Kök düğüm, tüm düğümlerin ebeveynidir ve adından da anlaşılacağı gibi, ağaçtaki en üstteki düğümdür. DT, her düğümün bir özelliği (nitelik), her bağlantının (dal) bir kararı (kural) gösterdiği ve her yaprağın bir sonucu (kategorik veya sürekli değer) gösterdiği bir yapıdır (Patel & Prajapati, 2018).
- *Rastgele Orman sınıflandırıcı,* karar ağaçlarının birleştirilmesi esasına dayalı bir regresyon ve sınıflandırma yöntemidir. Belirli bir örnek için ormandaki her ağaç üzerinde sınıflandırma yapılır. Orman daha sonra oylama işlemiyle örneğe ait sınıfını tespit eder. RF algoritmasında, karar ağaçlarının sonuçları birleştirilir ve orman adına tek bir karar verilir (Breiman, 2001).
- *K-En Yakın Komşu sınıflandırıcı*, basitliği ve nispeten yüksek yakınsama hızı nedeniyle sınıflandırma problemlerinde geniş bir yelpazede kullanılan bir yöntemdir. KNN, bir (x) örneğinden en yakın k örneği {i1, i2,.., ik} dikkate alır ve {c1, c2, ... ck} kümesindeki en sık sınıfa karar verir. En sık görülen sınıfın, o örneğin sınıfı olduğu varsayılır. En yakın örneği belirlemek için, KNN tekniği, saklanan örneklerin x ila k örneğinin yakınlığını ölçen bir mesafe metriğini benimser (Aldayel, 2012).

- *Gradient Boosting* algoritması, Regresyon ve sınıflandırma problemlerinde yüksek performans gösteren bir topluluk öğrenme algoritmasıdır. Algoritmada zayıf öğrenenleri iteratif olarak güçlü öğrenenlere dönüştürme amaçlanmaktadır. Gradyan artırmada üç öğe vardır. Bunlar kayıp fonksiyonu, zayıf öğrenen ve toplamsal modeldir (Nusrat ve ark., 2020).
- *XGBoost algoritması*, Chen ve Guestrin tarafından geliştirilen bir topluluk ağacı algoritmasıdır (Chen & Guestrin, 2016). Friedman'ın gradyan artırma (GB) algoritmasına dayalı olarak geliştirilmiştir (Friedman, 2001; Zhou ve ark., 2021). XGBoost, tahmin performansı tek başına kullanılan bireysel tekniklerden daha iyi olan, birleştirilmiş bir model üretmek için karar ağaçlarının verimli bir şekilde uygulanmasından oluşan kolektif bir modeldir (Jabeur ve ark., 2021).
- *Destek Vektör Makineleri*, denetimli bir regresyon ve sınıflandırma tekniğidir. SVM algoritmasında, her örneğin özellik vektörü, n'nin özellik sayısı olduğu n-boyutlu örnek uzayında bulunur. Daha sonra sınıflandırmayı sağlayan üst düzlem belirlenir ve sınıflandırma yapılır. En iyi üst düzlem, birbirine en yakın örnek noktalarından eşit uzaklıkta ve sınıfları en iyi ayıran düzlem olarak tanımlanır (Pisner & Schnyer, 2020).
- *Çok Katmanlı Algılayıcı*, insan beyninin çalışma şeklinden esinlenilerek geliştirilmiş yapıya sahip, ileri beslemeli bir yapay sinir ağı türüdür. Giriş katmanı, gizli katman ve çıkış katmanı olmak üzere üç katmandan oluşur. Bu yapıda, katmanlardaki düğümler sonraki katmandaki her bir düğüm ile tam bağlantılıdır (Potur & Erginel, 2021).

2.4. İlgili çalışmalar

Makine öğrenimi ve yapay zeka teknikleri, siber güvenlik alanında daha çok siber güvenlik farkındalık düzeylerinin tespitinde değil de zararlı yazılımların ve atakların tespitinde kullanılmıştır. Literatürde, SVM, NB, RF, Derin İnanç Ağları (Deep Belief Network: DBN), C4.5, Adaboost, Derin Sinir Ağları (Deep Neural Network: DNN) kullanarak *spam sınıflandırması* (Li ve ark., 2018; Ligthart ve ark., 2021; Makkar & Kumar, 2021; Srinivasan ve ark., 2021), Sinir Ağları (Neural Networks: NN), Derin Öğrenme (Deep Learning: DL), RF, SVM, NB, Lojistik Regresyon (Logistic Regression: LR), XGBoost, Karesel Ayırma Analizi (Quadratic Discriminant Analysis: QDA), KNN, C4.5, Bayesçi İnanç Ağları (Bayesian Belief Network: BBN) kullanarak *dolandırıcılık tespiti* (Xuan ve ark., 2018; Makki ve ark., 2019; Mittal & Tyagi, 2019), DNN, DenseNet, CNN, Bayes Ağı (Bayesian network: BN), KNN, MLP, J48 kullanarak *kötü amaçlı yazılım tespiti* (Narudin ve ark., 2016; Venkatraman ve ark., 2019; Xue ve ark., 2019; Gibert ve ark., 2020), NB, KNN, Adaboost, DT, SVM, LR, NN, XGBoost kullanarak *oltalama saldırısı* tespiti (Sahingoz ve ark., 2019; Shahrivari ve ark., 2020) gibi siber güvenlik alanlarında çokça kullanılmıştır.

Farklı odak ve metodolojilere sahip, siber güvenlik farkındalığı ve davranış olgusunu ele alan çalışmalar daha çok istatistiksel yöntemler üzerinde durmuşlardır. Bu istatistiksel çalışmaların ortak noktası, siber güvenlik farkındalığını, siber güvenlik davranışları etkileyen çeşitli faktörleri tanımlamaları ve bu faktörlerin birbiriyle bağlantılılığını açıklamaya çalışmalarıdır. Çalışmalar geliştirilmiş çeşitli siber güvenlik ölçeklerini kullanmışlardır. Bu çalışmalardaki hedef kitle ise çocuklardan başlayarak farklı eğitim düzeylerindeki öğrenciler, bilgi işlem personelleri gibi geniş bir yelpazede ele alınmıştır (Muhirwe & White, 2016; Karacı ve ark., 2017; Subramaniam, 2017; Özbek, 2019; Yiğit & Seferoğlu, 2019; Karakaya & Yetgin, 2020; Kovačević ve ark., 2020; Quayyum ve ark., 2021; Gündüzalp, 2021). Siber güvenlik kavramıyla önemli oranda benzerlikleri olan bilgi güvenliği farkındalığı ile ilgili Saridewi ve Sari (2021) Endonezya'da 488 kişiye anket uygulayarak bir çalışma gerçekleştirmişlerdir. Yaptıkları çalışmada SVM, KNN, LR, RF, DT, NB sınıflandırma algoritmalarını kullanmışlar aynı zamanda K-Means ve DBSCAN kümeleme tekniklerini kullanmışlardır. Sınıflandırma algoritmaları ile 0.89-0.99 arasında değişen doğruluk elde etmişlerdir (Saridewi & Sari, 2021). Siber güvenlik ölçekleri ele alınarak makine öğrenme yöntemlerini kullanan çalışma, literatürde incelendiği kadarıyla, sayısal olarak çok azdır. Balan ve ark. (2018) yaptığı çalışmada siber güvenlikle ilgili çoktan seçmeli bir testi Amerika Birleşik Devletleri'nde yaşayan yetişkin internet kullanıcılarına uygulayıp aldığı cevapların doğruluğuna göre Logistic Model Tree (LMT) algoritmasını kullanmışlardır. Kimlik doğrulama farkındalığı, özel tarama bilinci, oltalama saldırısının farkındalığı,

internet erişilebilirliğinin farkındalığı konusunda 0.81-0.99 oranında doğruluk elde etmişlerdir (Balan ve ark., 2018). Khan ve ark. (2021) sosyal medya kullanıcılarının siber güvenlik farkındalığını araştıran araştırmasında kullandıkları ölçek ile çeşitli analizler yapmışlar ve beş farklı ML (LR, NB, DT, SVM, KNN) tekniği kullanmışlar ve 0.66-0.75 arasında doğruluk elde etmişlerdir (Khan ve ark., 2021).

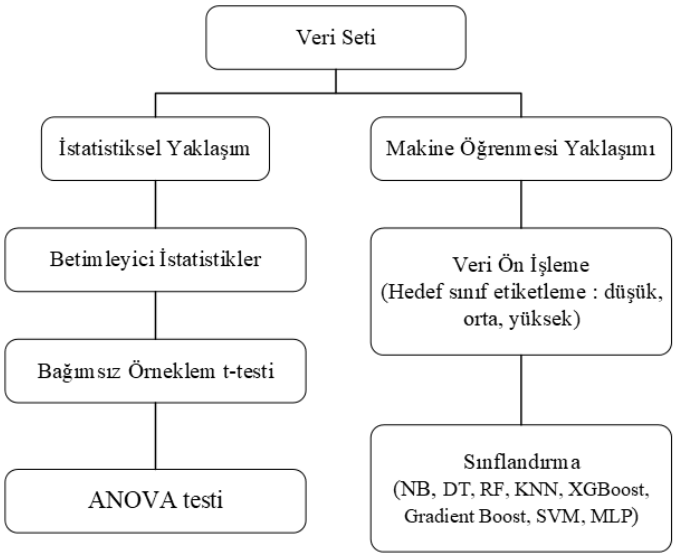
3. Yöntem

Siber güvenlik tehditleriyle ilgili riskleri elimine etmek veya minimum düzeye düşürmek için en önemli faktörlerden birisinin, siber güvenlik farkındalığı olduğu ifade edilmektedir (Safa ve ark., 2016). Bu bağlamda öğrencilerin siber güvenlik farkındalığı üzerinde durulması gereken bir konu olarak görülmüş ve çalışma bu nedenle gerçekleştirilmiştir.

Bu çalışmanın amacı, lisans öğrenimi gören öğrencilerin farklı değişkenlere göre siber güvenlik farkındalık seviyelerini belirlemektir. Bu amaca yönelik olarak aşağıdaki soruların cevapları aranmıştır:

- Kadın ve erkek öğrencilerin siber güvenlik farkındalık düzeyleri arasında anlamlı bir fark var mıdır?
- Öğrencilerin siber güvenlik farkındalık düzeyleri bölümlere göre değişmekte midir?
- Makine Öğrenme tekniklerinin, siber güvenlik farkındalık düzeylerini belirlemede performansı nedir?

Önerilen çalışma metodunda; ML algoritmalarına giriş olarak SGÖ'den elde edilen veriler kullanılmıştır. Çalışma iki farklı aşamadan oluşmaktadır. Öncelikle istatistiksel olarak betimsel tarama yöntemi benimsenmiş ardından ML teknikleri ile sınıflandırma yapılmıştır. Şekil 1'de çalışmanın yöntemi özetlenmiştir.



Şekil 1. Çalışma yöntemi.

3.1. Çalışma grubu

Araştırmanın evrenini Mehmet Akif Ersoy Üniversitesi, Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksekokulu bölümlerinde öğrenim gören öğrenciler oluşturmaktadır. Yüksekokulda 2021-2022 akademik yılı bahar döneminde kayıt yaptırıp, aktif olarak öğrenimine devam eden 734 öğrenci bulunmaktadır. Yüksekokulda 7+1 eğitim modeli uygulandığından çalışmanın yapıldığı bahar döneminde 4. Sınıf öğrencileri işletmede mesleki eğitim almaktadır. Bu nedenle işletmede mesleki eğitim alan 92 öğrenci araştırma kapsamı dışında tutulmuştur. Örneklem ise, Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksekokulu bölümlerinde öğrenim

gören 210 öğrenciden oluşmaktadır. Çalışmaya katılan öğrenci grubunun demografik özellikleri Çizelge 1'de verilmiştir.

Çizelge 1. Demografik özellikler

Bölümler	Cinsiyet		Toplam
	Erkek	Kadın	
Yönetim Bilişim Sistemleri	59	37	96
Muhasebe ve Finansal Yönetim	18	14	32
Gümrük İşletme	20	14	34
Bilişim Sistemleri ve Teknolojileri	22	2	24
Uluslararası Ticaret	12	12	24
Toplam	**131**	**79**	**210**

3.2. Veri toplama araçları

Veri toplamak için kullanılan ölçme aracı; cinsiyet, bölüm, sınıfa yönelik soruların olduğu kişisel bilgi bölümü ve Arpaci & Sevinç (2022) tarafından geliştirilen Siber Güvenlik Ölçeğinden (SGÖ) oluşmaktadır. Toplam 24 madde bulunan 5'li likert tipindeki ölçek, Gizlilik, Kontrol/Sahiplik, Bütünlük, Gerçeklik, Erişilebilirlik ve Fayda olmak üzere 6 faktörden oluşmaktadır. Ölçekteki ifadeler Kesinlikle Katılıyorum (5), Katılıyorum (4), Kararsızım (3), Katılmıyorum (2), Kesinlikle Katılmıyorum (1) şeklindedir. Ölçek puanlanırken; 17 normal madde ve 7 ters madde için farklı hesaplama yapılmıştır. Normal maddeler için puanlama "Kesinlikle Katılıyorum" seçeneğinden başlayarak 5'ten 1'e doğru yapılmış, ters maddeler için puanlama ise "Kesinlikle Katılmıyorum" seçeneğinden başlayarak 5'ten 1'e doğru yapılmıştır. SGÖ'nin tamamı için Cronbach Alfa iç güvenirlik katsayısı 0.88, Gizlilik faktörü için 0.784, Kontrol/Sahiplik faktörü için 0.810, Bütünlük faktörü için 0.795, Gerçeklik faktörü için 0.784, Erişilebilirlik faktörü için 0.734 ve Fayda faktörü için 0.735'dir. Bu çalışmada ise Ölçeğin Cronbach Alfa iç güvenirlik katsayısı 0.818 olarak bulunmuştur. Faktörlere göre ise sırasıyla 0.819, 0.912, 0.607, 0.826, 0.721 ve 0.714 olarak hesaplanmıştır. Elde edilen değerler orijinal ölçeğin değerlerine benzer değerlerdir. Hesaplanan değerlere göre ölçeğin tamamı ve 6 faktöre göre yapılan ölçümlerin yeterli güvenirliğe sahip olduğu söylenebilir.

3.3. Verilerin analizi

Veriler, dijital ortamda çevrimiçi olarak toplanmıştır. Verilerin analizi için yapılan istatistiksel analizlerde IBM SPSS yazılımı kullanılmıştır. Gerçekleştirilen analizler için 0.05 anlamlılık düzeyi kullanılmıştır. SGÖ ölçeğindeki maddelere ilişkin ortalama ve standart sapma değerleri hesaplanmış ve ölçekten elde edilen puanların normal dağılım gösterip göstermediğini tespit etmek amacıyla normallik testi yapılmış ve sonuçlar incelenmiştir. İncelemede çarpıklık (Skewness) ve basıklık (Kurtosis) katsayıları esas alınmıştır. Çizelge 2'de gösterilen çarpıklık ve basıklık katsayılarının kabul değerleri olan -1.5 ile +1.5 arasında olması nedeniyle verilerin normal dağıldığı kabul edilmiştir (Tabachnick & Fidell, 2013). Buna göre veri analizinde öğrencilerin cinsiyetleri bakımından farkın ortaya konulması amacıyla bağımsız örneklem t-testi, bölümlere göre gruplar arasındaki farkın tespiti için tek yönlü varyans analizi (ANOVA) testi kullanılmıştır. Ayrıca gruplar arası farkın kaynağını belirlemek amacıyla Games-Howell testi uygulanmıştır.

ML ile sınıflandırma için kurulan modellerin tasarımı Python programlama dili kullanılarak gerçekleştirilmiştir. Her model üzerindeki tüm deneyler, 10 katlı çapraz doğrulama kullanılarak gerçekleştirilmiştir. Bu değerlendirme, eğitim için 9 veriyi ve test için 1 veriyi ayırmaktadır. ML için ön hazırlık dosya yükleme, performans metriklerinin hesaplanmasında veri görselleştirme gibi işlemlerde Python numpy, sklearn, pandas, statistics, matplotlib.pyplot kütüphaneleri kullanılmıştır. Önerilen ML mimarilerinden RF, NB, SVM, DT, NN, Gradient Boost, MLP sınıflandırıcıları için sklearn kütüphanesi, XGBoost sınıflandırıcı için xgboost kütüphanesi kullanılmıştır. Çalışmada kullanılan ML yöntemlerine ait deneysel sonuçlar elde edilirken kullanılan model parametreleri Çizelge 2'de verilmiştir. KNN ve NB algoritmalarında ise "default" parametreler benimsenmiştir. Bu

parametreler belirlenirken veri sayısı, öznitelik sayısı, çıkış sayısı göz önünde bulundurularak çeşitli denemeler sonucunda belirlenmiştir.

Çizelge 2. Model hiperparametreleri

Model Adı	**Parametreler**
SVM	kernel=poly, degree=3, gamma="auto", cache_size=40, C=1
DT	max_depth=5
RF	max_depth=9, n_estimators=100, max_features=1
Gradient Boost	n_estimators=100, learning_rate=1.0, max_depth=1, random_state=0
XGBoost	max_depth=3, n_estimators=100, learning_rate=1, random_state=0, booster=gbtree
MLP	hidden_layer_sizes=5, max_iter=250, alpha=1e-4, solver="sgd" random_state=1, n_iter_no_change=50, learning_rate_init=0.01, momentum=0.3

Oluşturulan modellerin performansını değerlendirmek için kullanılan; doğruluk değeri Denklem 1'de, kesinlik (precision) değeri Denklem 2'de, duyarlılık (recall) değeri Denklem 3'te, F1-skoru ise Denklem 4'te gösterilmiştir.

(1)

(2)

(3)

(4)

4. Bulgular

Çalışmada kullanılan SGÖ'den elde edilen betimleyici istatistik değerleri Çizelge 3'te verilmiştir.

Çizelge 3. Betimleyici istatistikler

Faktörler	**N**	**Minimum**	**Maksimum**	**Ortalama**	**Ss**	**Çarpıklık**	**Basıklık**
Gizlilik	210	1.67	5	3.8571	1.08428	-1.084	.491
Kontrol/Sahiplik	210	2.4	5	4.0933	1.13915	-1.502	1.216
Bütünlük	210	1	5	3.1583	.85957	-.402	.598
Gerçeklik	210	1	5	2.1162	1.01010	1.104	.698
Erişilebilirlik	210	1	5	3.1655	1.13278	-.310	-.812
Fayda	210	1	5	3.5317	1.01624	-.736	.305

Çizelge 4'e göre öğrencilerin faktörlere göre aldığı puanların genel ortalaması 3.27 olarak bulunduğundan öğrencilerin siber güvenlik farkındalık düzeylerinin orta seviyede olduğu söylenebilir.

Veri analizinde öğrencilerin cinsiyetleri açısından anlamlı bir farklılık olup olmadığının ortaya konulması adına bağımsız örneklem t-testi yapılmış ve Çizelge 4'te faktörlere ait elde edilen değerler gösterilmiştir.

Çizelge 4. Öğrencilerin siber güvenlik farkındalıkları bağımsız örneklem t-testi sonuçları

Faktörler	Cinsiyet	N	X	Ss	t	sd	p
Gizlilik	Erkek	131	3.8015	1.16265	-.957	208	.340
	Kadın	79	3.9494	.93992			
Kontrol/Sahiplik	Erkek	131	3.9786	1.24165	-1.891	208	.060
	Kadın	79	4.2835	.92132			
Bütünlük	Erkek	131	3.0458	.92497	-2.473	208	.014
	Kadın	79	3.3449	.70518			
Gerçeklik	Erkek	131	2.1588	1.03796	.786	208	.433
	Kadın	79	2.0456	.96447			
Erişilebilirlik	Erkek	131	3.1870	1.24589	.354	208	.724
	Kadın	79	3.1297	.92149			
Fayda	Erkek	131	3.5776	1.05800	.842	208	.401
	Kadın	79	3.4557	.94459			

Çizelge 4'e göre öğrencilerin cinsiyetleri açısından siber güvenlik farkındalığında istatistiki olarak Gizlilik, Kontrol/Sahiplik, Gerçeklik, Erişilebilirlik, Fayda faktörlerinde anlamlı bir farklılık oluşturmadığı değerlendirilmiştir ($t = -1.891 - 0.842$; $p > .05$). Bütünlük faktörüne göre ise öğrencilerin cinsiyetleri açısından siber güvenlik farkındalığında anlamlı bir farklılık tespit edilmiştir ($t = -2.473$; $p < .05$).

Çalışmaya katılan öğrencilerin bölüm değişkeni Yönetim Bilişim Sistemleri, Muhasebe ve Finansal Yönetim, Gümrük İşletme, Bilişim Sistemleri ve Teknolojileri, Uluslararası Ticaret olarak ayarlanmıştır. Bölüm değişkeni ile ilgili olarak Çizelge 5'te, öğrencilerin bölümlerine göre grupların homojen dağılıp dağılmadığını belirlemek için yapılan homojenlik testi sonuçları, Çizelge 6'da ise gruplar arası farklılaşmayı tespit etmek için yapılan ANOVA testi verilmiştir.

Çizelge 5. Öğrencilerin bölümleri için yapılan homojenlik testi

Levene İstatistiği	df1	df2	p
13.744	4	205	.001

Grupların homojenliğini belirlemek için yapılan Levene testi sonucu p değeri 0.001<0.05 olarak bulunmuş ve bölümlere göre gruplar arası varyansların homojenliği testinde varyansların homojen olmadığı değerlendirilmiştir.

Çizelge 6. Bölümlere göre öğrencilerin siber güvenlik farkındalıkları ANOVA testi sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p
Gruplar arası	16.21	4	4.053	8.315	.001
Gruplar içi	99.911	205	0.487		
Toplam	116.121	209			

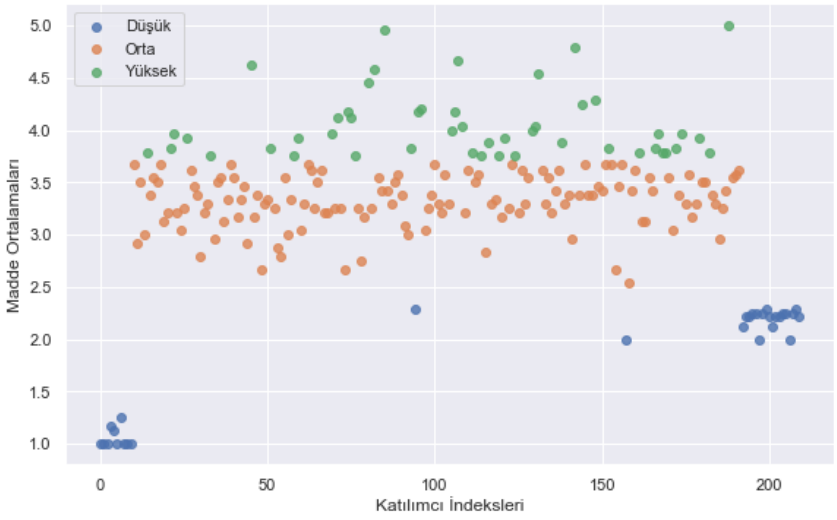
Çizelge 6'da verilen ANOVA testi sonucu, gruplar arası gözlenen farklılığın istatistiksel olarak anlamlı olduğu ortaya konulmuştur ($F(4\text{-}205)=8.314$, $p<.05$). Bu farkın grupların hangileri arasında olduğunun saptanması amacıyla Games-Howell testi yapılmıştır.

Çizelge 7'de gruplar arası karşılaştırmalar yapılmış ve karşılaştırılan bu grupların ortalama puanları arasındaki farklarına bakılmıştır. Bu değerlere bakıldığında Yönetim Bilişim Sistemleri-Muhasebe ve Finansal Yönetim, Bilişim Sistemleri ve Teknolojileri- Muhasebe ve Finansal Yönetim, Bilişim Sistemleri ve Teknolojileri-Uluslararası Ticaret bölümleri arasında anlamlı farklılıklar olduğu görülmüştür. Bu da Muhasebe ve Finansal Yönetim bölümünün siber güvenlik farkındalık puanlarının düşük olduğunun bir göstergesidir.

Çizelge 7. Öğrencilerin bölümleri arasındaki farklılığın anlamlılığına ilişkin Games-Howell sonuçları

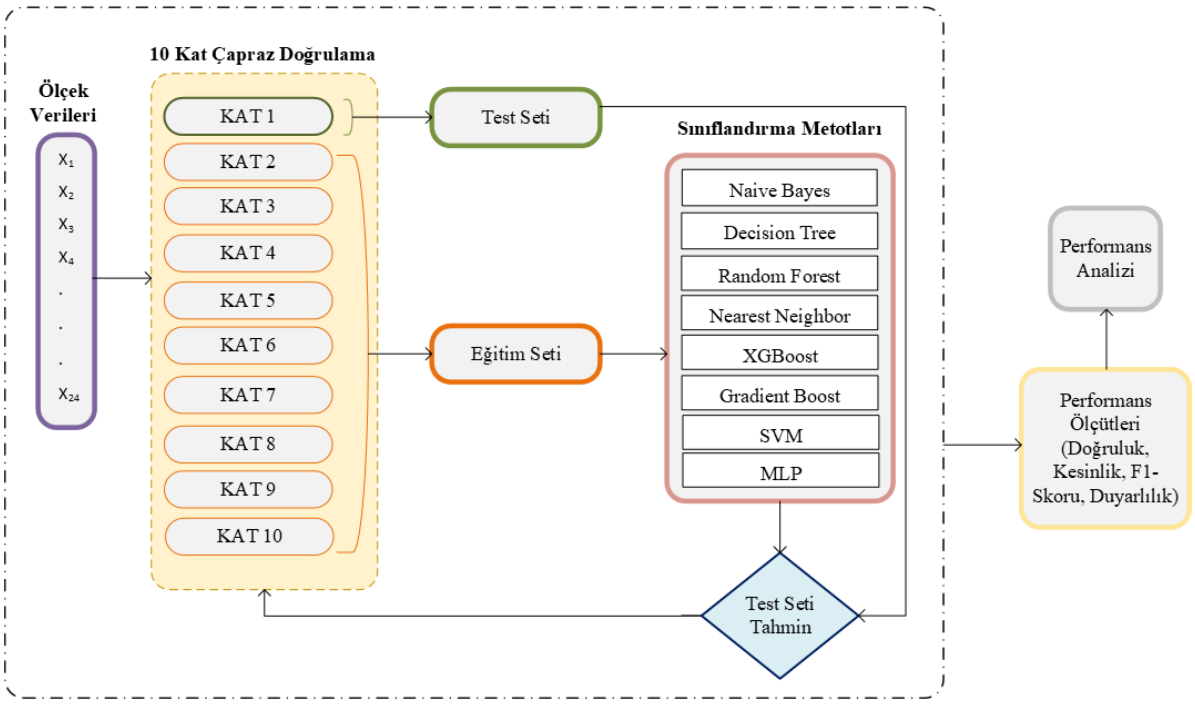
Bölüm	Bölüm	Ortalama Fark	Ss	p
Yönetim Bilişim Sistemleri	Muhasebe ve Finansal Yönetim	.73655*	0.19916	0.006
	Gümrük İşletme	0.04177	0.15276	0.999
	Bilişim Sistemleri ve Teknolojileri	-0.16319	0.09305	0.41
	Uluslararası Ticaret	0.27257	0.13395	0.272
Muhasebe ve Finansal Yönetim	Yönetim Bilişim Sistemleri	-.73655*	0.19916	0.006
	Gümrük İşletme	-.69478*	0.23733	0.038
	Bilişim Sistemleri ve Teknolojileri	-.89974*	0.20408	<.001
	Uluslararası Ticaret	-0.46398	0.22568	0.255
Gümrük İşletme	Yönetim Bilişim Sistemleri	-0.04177	0.15276	0.999
	Muhasebe ve Finansal Yönetim	.69478*	0.23733	0.038
	Bilişim Sistemleri ve Teknolojileri	-0.20496	0.15911	0.7
	Uluslararası Ticaret	0.2308	0.18601	0.728
Bilişim Sistemleri ve Teknolojileri	Yönetim Bilişim Sistemleri	0.16319	0.09305	0.41
	Muhasebe ve Finansal Yönetim	.89974*	0.20408	<.001
	Gümrük İşletme	0.20496	0.15911	0.7
	Uluslararası Ticaret	.43576*	0.14115	0.029
Uluslararası Ticaret	Yönetim Bilişim Sistemleri	-0.27257	0.13395	0.272
	Muhasebe ve Finansal Yönetim	0.46398	0.22568	0.255
	Gümrük İşletme	-0.2308	0.18601	0.728
	Bilişim Sistemleri ve Teknolojileri	-.43576*	0.14115	0.029

ML tekniklerinin, siber güvenlik farkındalık düzeylerini belirlemede öncelikle ölçek verileri eğitim ve test için hazırlanmıştır. Bu amaçla ölçeğe her katılan kişinin ölçekten aldığı toplam ortalama puan hesaplanmış ve bu değerler üç alt grupta kategorize edilmiştir: düşük (1-2.33), orta (2.34-3.67) ve yüksek (3.68-5) (D'Silva ve ark., 2010; Hassan ve ark., 2011; Ramli ve ark., 2013). Ankete katılan her öğrencinin verdiği cevaplar puanlanmış ve kategorize edilen değerlere göre, öğrencinin siber güvenlik seviyesi "düşük", "orta", "yüksek" olarak etiketlenmiştir. Kategorize edilen değerler Şekil 2'de gösterilmiştir.



Şekil 2. Veri seti dağılımı.

Ölçekteki her bir soru, tasarlanan sınıflandırıcı için giriş olarak verilmiştir. Sınıflandırıcının eğitimi ve testinde 10 katlı çapraz doğrulama kullanılmıştır. Sınıflandırmanın çıkışında ise öğrencilerin siber güvenlik farkındalık düzeyleri düşük, orta, yüksek olarak alınmış ve performans metrikleri elde edilmiştir. Önerilen ML yaklaşımı Şekil 3'te gösterilmiştir.

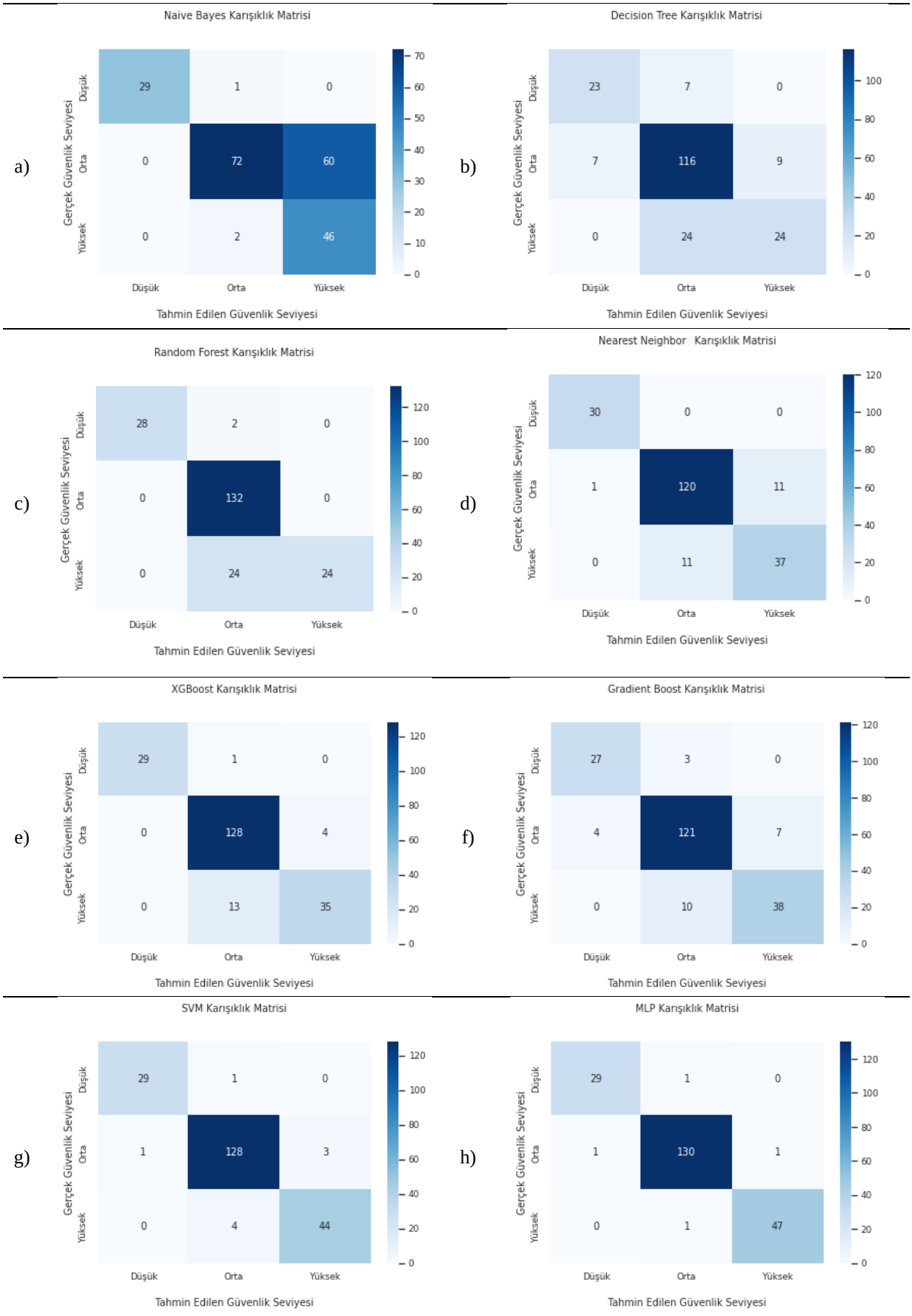


Şekil 3. Önerilen çalışmanın ML modeli.

Araştırmada kullanılan tüm modellere ait eğitim ve test sonuçlarına ait metrikler Çizelge 8'de gösterilmiştir. Yapılan testler sonucunda 0.7-0.98 arasında değişen doğruluk değerleri, 0.7-0.96 arasında değişen F1 skorları elde edilmiştir. Elde edilen test ve eğitim skorları 10 kat çapraz doğrulama sonucunda elde edilen ortalama değerlerdir. Çizelge 8'de ortalama skorların yanı sıra her bir algoritmaya ait test ve eğitim skorlarının Standart sapma değerleri verilmiştir. Şekil 4'te ise çalışmada kullanılan modellerin eğitilmesi sonucunda yapılan testler sonucunda elde edilen karışıklık matrisleri verilmiştir.

Çizelge 8. Eğitim ve test performans metrikleri

Algoritmalar		**Test**				**Eğitim**			
		Doğruluk	Kesinlik	F1 Skoru	Duyarlılık	Doğruluk	Kesinlik	F1-Skoru	Duyarlılık
NB	Ort.	0.7000	0.7914	0.7369	0.8143	0.7254	0.8170	0.7807	0.8520
	Ss	0.0641	0.0418	0.0580	0.0658	0.0143	0.0066	0.0111	0.0081
DT	Ort.	0.7905	0.7538	0.7048	0.7166	0.9122	0.9352	0.9048	0.8920
	Ss	0.1048	0.1568	0.1549	0.1498	0.0297	0.0276	0.0364	0.0431
RF	Ort.	0.8476	0.8867	0.7858	0.7713	1.0000	1.0000	1.0000	1.0000
	Ss.	0.0873	0.1043	0.0934	0.0919	0.0000	0.0000	0.0000	0.0000
Gradient Boost	Ort.	0.8857	0.8770	0.8597	0.8718	1.0000	1.0000	1.0000	1.0000
	Ss.	0.0713	0.1050	0.0907	0.0547	0.0000	0.0000	0.0000	0.0000
KNN	Ort.	0.8905	0.8478	0.8571	0.8832	0.9550	0.9544	0.9552	0.9563
	Ss.	0.0641	0.1047	0.0990	0.0894	0.0049	0.0073	0.0039	0.0044
XGBoost	Ort.	0.9143	0.9367	0.9025	0.8943	1.0000	1.0000	1.0000	1.0000
	Ss.	0.0732	0.0428	0.0634	0.0671	0.0000	0.0000	0.0000	0.0000
SVM	Ort.	0.9571	0.9513	0.9324	0.9362	1.0000	1.0000	1.0000	1.0000
	Ss	0.0256	0.0523	0.0589	0.0621	0.0000	0.0000	0.0000	0.0000
MLP	Ort.	0.9810	0.9711	0.9668	0.9749	1.0000	1.0000	1.0000	1.0000
	Ss	0.0233	0.0504	0.0482	0.0490	0.0000	0.0000	0.0000	0.0000



Şekil 4. Karışıklık matrisleri a: NB, b: DT, c: RF, d: KNN, e: XGBoost, f: Gradient Boost, g: SVM, h: MLP.

5. Tartışma ve Sonuç

Bu çalışmada, Mehmet Akif Ersoy Üniversitesi, Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksekokulu bölümlerinde öğrenim gören öğrencilerin siber güvenlik farkındalık düzeyleri makine öğrenme yöntemleriyle belirlenmiştir. ML algoritmalarına giriş olarak SGÖ verileri kullanıldığından çalışmada öncelikle istatistiksel olarak betimsel tarama yöntemi benimsenmiş ardından ML teknikleri ile sınıflandırma yapılmıştır.

SGÖ'den elde edilen ortalama puanlar incelendiğinde, öğrencilerin siber güvenliğe yönelik farkındalıklarının genel itibarıyla orta düzeyde olduğu görülmektedir. Ancak ölçeğin gerçeklik faktöründeki ortalama puanların düşük olduğu gözlemlenmiştir. Buradaki puan düşüklüğünün, spesifik olarak oltalama saldırılarının nasıl gerçekleştiği ile ilgili öğrencilerin yeterli bilgi birikimine sahip olmamalarından kaynaklandığı düşünülmektedir. Çalışma sonuçlarına göre erkek öğrenciler ile kadın öğrencilerin siber güvenlik farkındalıkları bakımından 5 faktöre göre anlamlı bir farklılık tespit edilememiştir. Ancak Bütünlük faktöründe siber güvenlik farkındalıkları bakımından anlamlı bir farklılık görülmüştür. Kadın öğrencilerin siber ortamda saklanan verilerin güvenliğine daha şüpheci bir bakış açısı gösterdiği düşünülmektedir. Öğrencilerin bölümleri arasında siber güvenlik farkındalıkları bakımından anlamlı farklılıklar olduğu görülmüştür. Bilişim sistemleri ile ilgili bölümlerin, diğer bölümlere göre siber güvenlik farkındalıklarının daha yüksek olduğu görülmüştür. Bunun sebebi ise lisans eğitimi sırasında alınan derslerin ve içeriklerinin etkisi olarak değerlendirilmiştir.

Öğrencilerin ölçekten aldıkları puanlar düşük, orta ve yüksek olarak etiketlenmiş ve ölçek soruları RF, NB, SVM, DT, NN, XGBoost, Gradient Boost, MLP algoritmalarına giriş olarak verilmiştir. Daha sonra kurulan her bir model eğitilip test edildikten sonra her bir öğrencinin siber güvenlik farkındalık düzeyleri tespit edilmiştir.

Düşük, orta, yüksek olarak etiketleme, düşük 1-2.33, orta 2.34-3.67 ve yüksek 3.68-5 puan ortalaması aralığına göre yapılmıştır. Çalışmada kullanılan sınıflandırma yöntemlerindeki karışıklık matrisleri ve performans metrikleri incelendiğinde; NB, DT, RF, KNN, XGBoost, Gradient Boost algoritmalarının orta ve yüksek etiketli verilerin sınıflandırılmasında başarısız kaldığı gözlemlenmiştir. Elde edilen performans metriklerine göre NB algoritması ile 0.70, DT algoritması ile 0.7905, RF algoritması ile 0.8476, Gradient Boost algoritması ile 0.8857, KNN algoritması ile 0.8905, XGBoost algoritması ile 0.9143, SVM algoritması ile 0.9571, MLP algoritması ile 0.981 doğruluk değerleri elde edilmiştir. MLP ve SVM algoritmalarının hem doğruluk hem de F1-skoruna bakıldığında diğer algoritmalara göre performansının yüksek olduğu görülmüştür. Siber güvenlik farkındalık seviyelerinin belirlenmesinde ML tekniklerinin kullanılabileceği, geliştirilen veya geliştirilecek olan farklı ML teknikleriyle performans metriklerinin daha yüksek skorlara çıkarılabileceği öngörülmektedir.

Etik Kurul Onayı

Bu çalışmada ulusal ve uluslararası etik kurallara uyulmuştur. Bu çalışma, Mehmet Akif Ersoy Üniversitesi, Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu (GO 2022/824) tarafından etik açıdan onaylanmıştır.

Kaynakça

Abomhara, M., & Køien, G. M. (2015). Cyber Security and the Internet of Things: Vulnerabilities, Threats, Intruders and Attacks. *Journal of Cyber Security and Mobility*, *4*(1), 65-88.

Aldayel, M. S. (2012, December). *K-Nearest Neighbor classification for glass identification problem*. 2012 International Conference on Computer Systems and Industrial Informatics, Sharjah, United Arab Emirates. doi:10.1109/ICCSII.2012.6454522

Alzahrani, L. (2021). Statistical analysis of cybersecurity awareness issues in higher education institutes. *International Journal of Advanced Computer Science and Applications*, 12(11), 630-637. doi:10.14569/IJACSA.2021.0121172

Arpaci, I., & Sevinc, K. (2022). Development of the cybersecurity scale (CS-S): Evidence of validity and reliability. *Information Development*, 38(2), 218-226. doi:10.1177/0266666921997512

Balan, S., Gawand, S., & Purushu, P. (2018). Application of machine learning classification algorithm to cybersecurity awareness. *Information Technology & Management Science (RTU Publishing House)*, 21, 45-48. doi:10.7250/itms-2018-0006

Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (pp. 403-412). Amsterdam, The Netherlands: Elsevier Science Publisher. doi:10.1016/B978-0-12-809633-8.20473-1

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32. doi:10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016, August). *Xgboost: A scalable tree boosting system*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California, USA. doi:10.1145/2939672.2939785

D'Silva, J. L., Samah, B. A., Shaffril, H. A. M., & Hassan, M. A. (2010). Factors that influence attitude towards ICT usage among rural community leaders in Malaysia. *Australian Journal of Basic and Applied Sciences*, 4(10), 5214-5220.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2012). *How to design and evaluate research in education* (C. 7). New York, USA: McGraw-hill Education.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232. doi:10.1214/aos/1013203451

Gibert, D., Mateu, C., & Planes, J. (2020). HYDRA: A multimodal deep learning framework for malware classification. *Computers & Security*, 95, 101873. doi:10.1016/j.cose.2020.101873

Gündüzalp, C. (2021). Üniversite çalışanlarının dijital veri ve kişisel siber güvenlik farkındalıkları (bilgi işlem daire başkanlıkları örneği). *Journal of Computer and Education Research*, 9(18), 598-625. doi:10.18009/jcer.907022

Hassan, M. A., Samah, B. A., Shaffril, H. M., & D'Silva, J. L. (2011). Perceived usefulness of ICT usage among JKKK members in Peninsular Malaysia. *Asian Social Science*, 7(10). doi:10.5539/ass.v7n10p255

Humayun, M., Niazi, M., Jhanjhi, N. Z., Alshayeb, M., & Mahmood, S. (2020). Cyber security threats and vulnerabilities: A systematic mapping study. *Arabian Journal for Science and Engineering*, 45(4), 3171-3189. doi:10.1007/s13369-019-04319-2

IWS. (2022). Internet World Stats. https://www.internetworldstats.com/europa2.htm#tr Erişim Tarihi: 18 Ağustos 2022.

İlker, K. (2019). Kaba kuvvet saldırı tespiti ve teknik analizi. *Sakarya University Journal of Computer and Information Sciences*, 2(2), 61-69. doi:10.35377/saucis.02.02.561844

Jabeur, S. B., Mefteh-Wali, S., & Viviani, J.-L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 1-21. doi:10.1007/s10479-021-04187-w

Karacı, A., Akyüz, H. İ., & Bilgici, G. (2017). Üniversite öğrencilerinin siber güvenlik davranışlarının incelenmesi. *Kastamonu Eğitim Dergisi*, 25(6), 2079-2094. doi:10.24106/kefdergi.351517

Karakaya, A., & Yetgin, M. A. (2020). Karabük üniversitesi çalışanlarına yönelik kişisel siber güvenlik üzerine araştırma. *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 10(2), 157-172. doi:10.47147/ksuiibf.816171

Khan, F., Ncube, C., Ramasamy, L. K., Kadry, S., & Nam, Y. (2020). A digital DNA sequencing engine for ransomware detection using machine learning. *IEEE Access*, 8, 119710-119719. doi:10.1109/ACCESS.2020.3003785

Khan, N. F., Ikram, N., Murtaza, H., & Asadi, M. A. (2021). Social media users and cybersecurity awareness: Predicting self-disclosure using a hybrid artificial intelligence approach. *Kybernetes*, 52(1), 401-421. doi:10.1108/K-05-2021-0377

Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. *IEEE Communications Surveys & Tutorials*, 15(4), 2091-2121. doi:10.1109/SURV.2013.032213.00009

Kovačević, A., Putnik, N., & Tošković, O. (2020). Factors related to cyber security behavior. *IEEE Access*, 8, 125140-125148. doi:10.1109/ACCESS.2020.3007867

Li, Y., Nie, X., & Huang, R. (2018). Web spam classification method based on deep belief networks. *Expert Systems with Applications*, 96, 261-270. doi:10.1016/j.eswa.2017.12.016

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification. *Applied Soft Computing*, 101, 107023. doi:10.1016/j.asoc.2020.107023

Makkar, A., & Kumar, N. (2021). PROTECTOR: An optimized deep learning-based framework for image spam detection and prevention. *Future Generation Computer Systems*, 125, 41-58. doi:10.1016/j.future.2021.06.026

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M.-S., & Zeineddine, H. (2019). An experimental study with imbalanced classification approaches for credit card fraud detection. *IEEE Access*, 7, 93010-93022. doi:10.1109/ACCESS.2019.2927266

Mittal, S., & Tyagi, S. (2019, January). *Performance evaluation of machine learning algorithms for credit card fraud detection*. 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). doi:10.1109/CONFLUENCE.2019.8776925

Muhirwe, J., & White, N. (2016). Cybersecurity awareness and practice of next generation corporate technology users. *Issues in Information Systems*, 17(2), 183-192. doi:10.48009/2_iis_2016_183-192

Narudin, F. A., Feizollah, A., Anuar, N. B., & Gani, A. (2016). Evaluation of machine learning classifiers for mobile malware detection. *Soft Computing*, 20(1), 343-357. doi:10.1007/s00500-014-1511-6

Nusrat, F., Uzbaş, B., & Baykan, Ö. K. (2020). Prediction of diabetes mellitus by using gradient boosting classification. *Avrupa Bilim ve Teknoloji Dergisi*, Ejosat Special Issue 2020, 268-272. https://doi.org/10.31590/ejosat.803504

Özbek, Y. (2019). *Öğretmen adaylarının siber güvenlik farkındalıklarının incelenmesi*. (Doktora Tezi), Necmettin Erbakan Üniversitesi, Eğitim Bilimleri Enstitüsü, Konya, Türkiye.

Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74-78. doi:10.26438/ijcse/v6i10.7478

Pisner, D. A., & Schnyer, D. M. (2020). Chapter 6 - Support Vector Machine. A. Mechelli & S. Vieira (Ed.), *Machine Learning* (pp. 101-121). Cambridge, USA: Academic Press. doi:10.1016/B978-0-12-815739-8.00006-7

Potur, E. A., & Erginel, N. (2021). Kalp yetmezliği hastalarının sağ kalımlarının sınıflandırma algoritmaları ile tahmin edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (24), 112-118. doi:10.31590/ejosat.902357

Quayyum, F., Cruzes, D. S., & Jaccheri, L. (2021). Cybersecurity awareness for children: A systematic literature review. *International Journal of Child-Computer Interaction*, 30, 100343. doi:10.1016/j.ijcci.2021.100343

Ramli, S. A. B., Omar, S. Z., Bolong, J., D'Silva, J. L., & Shaffril, H. A. M. (2013). Influence of behavioral factors on mobile phone usage among fishermen: The case of Pangkor Island Fishermen. *Asian Social Science*, 9(5), 162. doi:10.5539/ass.v9n5p162

Safa, N. S., Von Solms, R., & Futcher, L. (2016). Human aspects of information security in organisations. *Computer Fraud & Security*, 2016(2), 15-18. doi:10.1016/S1361-3723(16)30017-3

Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications*, 117, 345-357. doi:10.1016/j.eswa.2018.09.029

Saridewi, V. S., & Sari, R. F. (2021). Implementation of machine learning for human aspect in information security awareness. *Journal of Applied Engineering Science*, 19(4), 1126-1142. doi:10.5937/jaes0-28530

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data*, 7, 41. doi:10.1186/s40537-020-00318-5

Shahrivari, V., Darabi, M. M., & Izadi, M. (2020). Phishing detection using machine learning techniques. *arXiv preprint arXiv:2009.11116*. doi:10.48550/arXiv.2009.11116

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. *Energies*, 13(10), 2509. doi:10.3390/en13102509

Srinivasan, S., Ravi, V., Alazab, M., Ketha, S., Al-Zoubi, A., & Padannayil, S. K. (2021). Spam emails detection based on distributed word embedding with deep learning. In Y. Maleh, M. Shojafar, M. Alazab, & Y. Baddi (Eds.), *Machine Intelligence and Big Data Analytics for Cybersecurity Applications* (pp. 161-189). Switzerland: Springer Cham.

Subramaniam, S. R. (2017, December). *Cyber security awareness among Malaysian pre-university students*. Proceeding of the 6th Global Summit on Education, Kualalumpur, Malasia.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics*. Boston, USA: Pearson Education.

TÜİK. (2021). Türkiye İstatistik Kurumu. https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2021-37437 Erişim Tarihi: 18 Ağustos 2022.

Venkatraman, S., Alazab, M., & Vinayakumar, R. (2019). A hybrid deep learning image-based analysis for effective malware detection. *Journal of Information Security and Applications*, 47, 377-389. doi:10.1016/j.jisa.2019.06.006

Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *Computers & Security*, 38, 97-102. doi:10.1016/j.cose.2013.04.004

Weamie, S. J. (2022). Cross-site scripting attacks and defensive techniques: A comprehensive survey. *International Journal of Communications, Network and System Sciences*, 15(8), 126-148. doi:10.4236/ijcns.2022.158010

Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018, March). *Random forest for credit card fraud detection*. 2018 IEEE 15th international conference on networking, sensing and control (ICNSC), Zhuhai, China. doi:10.1109/ICNSC.2018.8361343

Xue, D., Li, J., Lv, T., Wu, W., & Wang, J. (2019). Malware classification using probability scoring and machine learning. *IEEE Access*, 7, 91641-91656. doi:10.1109/ACCESS.2019.2927552

Yalçınkaya, M. A., & Küçüksille, E. (2021). Web uygulama sızma testlerinde kapsam genişletme işlemi için metodoloji geliştirilmesi ve uygulanması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *25*(1), 16-27. doi:10.19113/sdufenbed.661867

Yiğit, M. F., & Seferoğlu, S. S. (2019). Öğrencilerin siber güvenlik davranışlarının beş faktör kişilik özellikleri ve çeşitli diğer değişkenlere göre incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 15(1), 186-215. doi:10.17860/mersinefd.437610

Zhou, J., Qiu, Y., Khandelwal, M., Zhu, S., & Zhang, X. (2021). Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *International Journal of Rock Mechanics and Mining Sciences*, 145, 104856. doi:10.1016/j.ijrmms.2021.104856