YTÜ FEN EDEBİYAT FAKÜLTESİ

MATEMATİK BÖLÜMÜ

SİBER GÜVENLİKTE MAKİNE ÖĞRENMESİ

Elifnur TOMAZ 16025032

Danışman: Doç. Dr. Muttalip ÖZAVŞAR

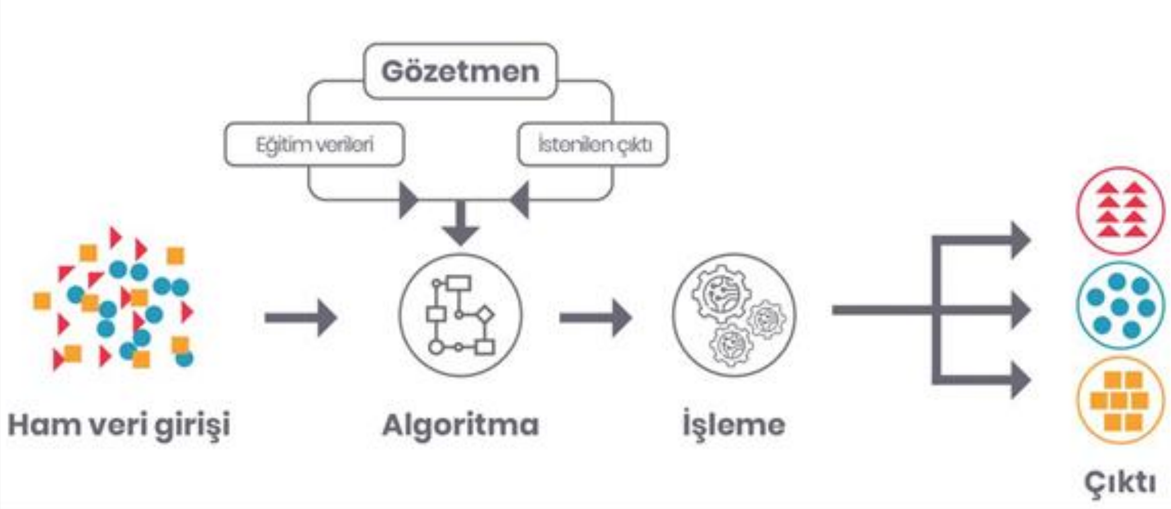
ÖZET

Her alana yayılarak hayatımızı kolaylaştıran yapay zekanın bir dalı olan makine öğrenmesinin, sanal bilginin güvenliğinin çok önemli olduğu şu zamanlarda siber güvenlikte kullanım amacına, sağladığı kolaylıklar ile birlikte oluşabilecek zorluklarına değinerek bu alanda kullanılan araçları inceledik.

MAKİNE ÖĞRENMESİ (ML)

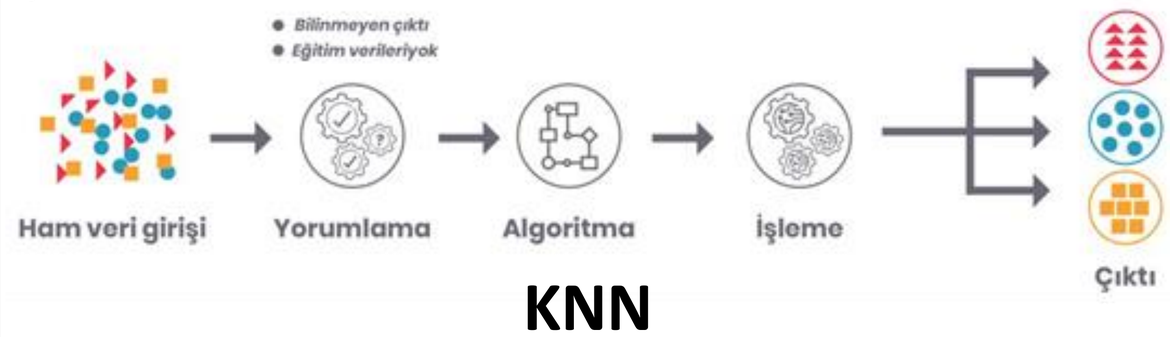
Denetimli Öğrenme (Supervised Learning)

Denetimli makine öğrenimi algoritmaları en yaygın olarak kullanılanlardır. Bu model sayesinde, veri uzmanı bir kılavuz olarak hareket eder ve algoritmaya hangi sonuçlara varması gerektiğini öğretir. Örneğin; malwarelerin tespit edilmesinde ve sınıflandırmasında bu yöntemler kullanılır.



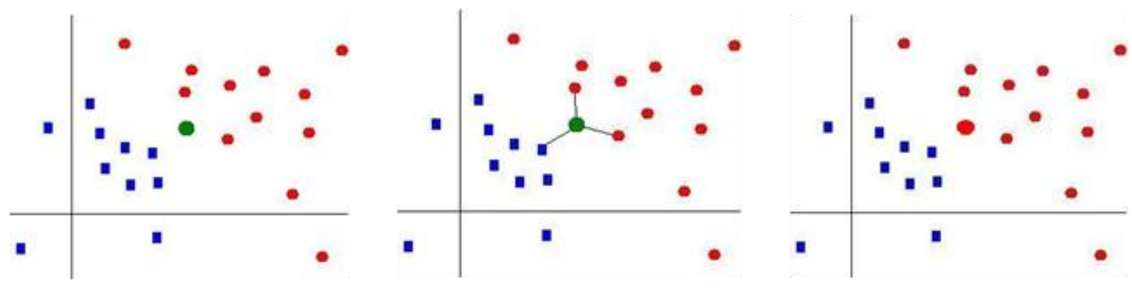
Denetimsiz Öğrenme (Unsupervised Learning)

Denetlenmeyen makine öğrenimi, bir insan tarafından sürekli ve yakın kılavuzluk sağlanmadan bilgisayarın karmaşık süreçleri ve modelleri öğrendiği daha bağımsız bir yaklaşımdan yararlanır. Denetlenmeyen makine öğrenimi, etiketleri veya spesifik, tanımlanmış bir çıktısı olmayan verilere dayalı eğitimi içerir.



KNN

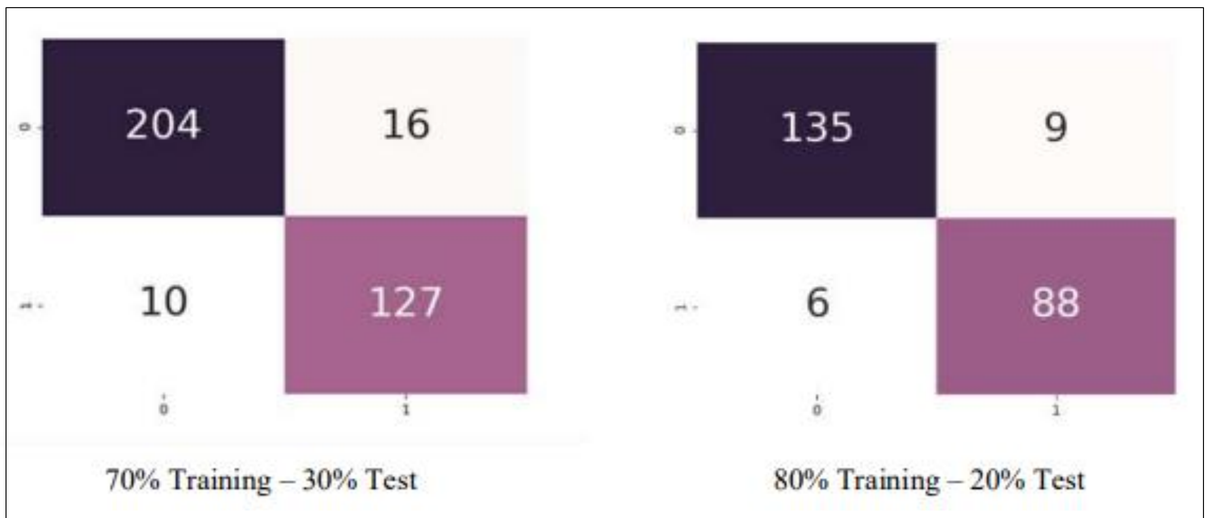
KNN (K-Nearest Neighbors) Algoritması bir supervised öğrenmesidir. gözlemlerin birbirlerine olan benzerlikleri üzerinden tahminlerin yapıldığı gözetimli makine öğrenmesi modellerinde regresyon ve sınıflandırma problemlerinde kullanılan bir algoritmadır.



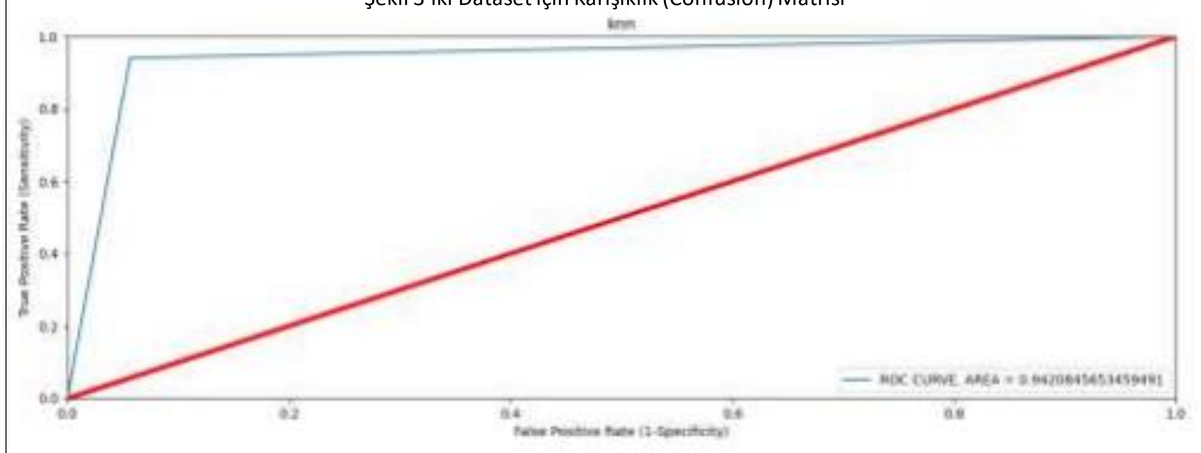
Malware tespiti için bir KNN Örneği

Eğitim ve test verilerinin seçiminde rastgele seçilen datasetler kullanılır. Buna göre test hem %70-%30 hem de %80-%20 eğitim ve test oranlarıyla gerçekleştirilir. Bu seçimler farklı deneyler için birçok kez tekrarlanmış ve sonuçlar bu bölümde açıklanmıştır. Performans değerlendirmesi için kesinlik, geri çağırma ve f ölçümü metrikleri kullanılır.

Mesafe metriği 5 olarak ayarlanmıştır. Öklid mesafesi, 5 örnek noktalı test verilerinin mesafesini hesaplamak için kullanılır. Mesafe ölçütü olarak Minkowski seçilmiştir. Toplamda 1189 mobil uygulamada yapılmıştır. Şekil 4.10'da verilen seçimler hem %70-%30 hem de %80-%20 için elde edilen karışıklık matrisleri görülmektedir.



Şekil 3 İki Dataset için Karışıklık (Confusion) Matrisi



Şekil 5 Modelin ROC Eğrisi

Önerilen model ile her iki testte de başarılı sınıflandırma sonuçları elde edilmiş olup, FP (False Positive) ve FN (False Negative) sayısı çok düşüktür. İki veri seçme yöntemiyle yapılan test sonuçlarının karşılaştırılması, başta sistem sınıflandırması olmak üzere tüm ölçüm metriklerinde daha güçlü sonuçlara olanak sağlar. Bu da daha kötü niyetli ve iyi huylu uygulamalar içeren veri kümeleri ile başarının daha yüksek düzeyde yakalanmasının mümkün olacağını göstermektedir. Modelin ROC olasılık eğrisi Şekil 5'da gösterilmektedir.

K value	Accuracy (%)	Recall (%)	Precision (%)	F-measure (%)
1	91.0	90.5	86.7	88.6
2	93.8	88.3	95.3	91.7
3	93.5	93.4	90.1	91.8
4	93.2	91.2	91.2	91.2
5	**94.1**	**91.2**	**92.7**	**92.4**
6	93.8	92.7	91.4	92.0
7	92.9	93.4	88.9	91.1
8	92.7	92.7	88.8	90.7
9	93.8	93.4	90.8	92.1
10	93.1	91.2	93.4	92.3

Şekil 6 K Değeri değişiminin model tanıma performansına etkisi

Şekil 6'da gösterilen sonuçlarda tanıma performansı hiçbir zaman %90'ın altına düşmedi. Bu durum önerilen modelin yüksek tanınma oranına sahip olduğunu göstermektedir. Parametrelerde yapılan değişiklikler sonuçların belirli bir seviyenin altına düşmesini etkilemez. Ancak en iyi seçim ile diğer modellere göre en iyi sonuçlar elde edilir. Sonuçlar, malware tespiti için KNN sınıflandırıcısının kullanılmasının etkisini göstermiştir.

KAYNAKÇA

[1] https://www.oracle.com/tr/artificial-intelligence/machine-learning/what-is-machine-learning/
[2] https://aws.amazon.com/tr/what-is/reinforcement-learning/
[3] https://www.turhost.com/blog/makine-ogrenmesi-machine-learning-nedir/
[4] https://miuul.com/not-defteri/k-en-yakin-komsu-algoritmasi-nasil-calisir
[5] https://bilgisayarkavramlari.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/
[8] A. H. Yurttakal, R. S. Arslan, H. Candan, "K-NEAREST NEIGHBOUR CLASSIFIER USAGE FOR PERMISSION BASED MALWARE DETECTION IN ANDROID", Iontech Journal, 15-27, 2020