T.C.

İSTANBUL KÜLTÜR ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ARAÇ İÇİ AĞLARDA MAKİNE ÖĞRENMESİ TABANLI SALDIRI TESPİT SİSTEMİ

(Machine Learning Based Intrusion Detection System for in-Vehicle Networks)

YÜKSEK LİSANS TEZİ

Soner Can KALKAN

1401020022

Anabilim Dalı: Bilgisayar Mühendisliği

Program: Bilgisayar Mühendisliği

Tez Danışmanı: Prof. Dr. Özgür Koray ŞAHİNGÖZ

NİSAN 2022

T.C.

İSTANBUL KÜLTÜR ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ARAÇ İÇİ AĞLARDA MAKİNE ÖĞRENMESİ TABANLI SALDIRI TESPİT SİSTEMİ

(Machine Learning based Intrusion Detection System for in-Vehicle Networks)

YÜKSEK LİSANS TEZİ

Soner Can KALKAN

1401020022

Anabilim Dalı: Bilgisayar Mühendisliği

Program: Bilgisayar Mühendisliği

Tez Danışmanı: Prof. Dr. Özgür Koray ŞAHİNGÖZ
Diğer Jüri Üyeleri: Prof. Dr. Murat ERMİŞ
Dr.Öğ.Üyesi İsmail KOÇ

NİSAN 2022

ÖNSÖZ

“Araç İçi Ağlarda Makine Öğrenimi Tabanlı Saldırı Tespit Sistemi (Machine Learning based Intrusion Detection System for in-Vehicle Networks)” adlı yüksek lisans tez çalışmam süresince bilgi ve deneyimi ile çalışmalarıma yön veren ve büyük bir sabırla desteğini esirgemeyen değerli tez danışmanım Prof. Dr. Özgür Koray Şahingöz’e teşekkürlerimi sunarım.

İÇİNDEKİLER

ŞEKİL LİSTESİ

TABLO LİSTESİ

KISALTMALAR

CAN	: Controller Area Network - Denetleyici Alan Ağı
ID	: Identification Number - Kimlik Değeri
IDS	: Intrusion Detection Sytem - Saldırı-Sızma Tespit Sistemi
RF	: Random Forest - Rastgele Orman
NB	: Naive Bayes - Naif Bayes
DT	: Decision Tree - Karar Ağacı
MLP	: Multi Layer Perceptron - Çok Katmanlı Algılayıcı
AA	: Adaboost - Adaptif Arttırma Sınıflandırması
CV	: Cross Validation - Çapraz Doğrulama
CM	: Confusion Matrix - Karmaşıklık Matrisi
PCA	: Principal Component Analysis - Temel Bileşen Analizi
RSU	: Roadside Unit - Yol Haberleşme Ünitesi
UGB	: Ultra Wide Band - Ultra Geniş Bant
DOS	: Denial Of Service - Kullanım Dışı Bırakma Saldırısı
KNN	: K'th Nearest Neighbor - K En Yakın Komşu
SVM	: Support Vector Machine - Destek Vektör Makinesi
GAN	: Generative Adversarial Netowork - Rekabetçi Üretici Ağ
V2V	: Vehicle To Vehicle - Araçtan Araca İletişim
V2P	: Vehicle To Person - Araç-Yaya İletişimi
V2N	: Vehicle To Network - Araç-Ağ İletişimi
V2U	: Vehicle To User - Araç-Kullanıcı İletişimi
OBD	: On-board Diagnostics - Araç İçi Hata Tespit Portu
ECU	: Engine Control Unit - Motor Kontrol Birimi
VRU	: Vulnerable Road User - Kırılgan Yol Kullanıcısı
LSTM	: Long Short Term Memory - Uzun Kısa Süreli Bellek

Üniversite	**:**	**T.C. İstanbul Kültür Üniversitesi**
Enstitüsü	**:**	**Lisansüstü Eğitim Enstitüsü**
Anabilim Dalı	**:**	**Bilgisayar Mühendisliği**
Program	**:**	**Bilgisayar Mühendisliği**
Tez Danışmanı	**:**	**Prof. Dr. Özgür Koray ŞAHİNGÖZ**
Tez Türü ve Tarihi	**:**	**Yüksek Lisans – Nisan 2022**

ÖZET

ARAÇ İÇİ AĞLARDA MAKİNE ÖĞRENİMİ TABANLI SALDIRI TESPİT SİSTEMİ

Gelişen dünyaya paralel olarak ulaşım teknolojileri de her geçen yıl önemli ölçüde gelişmeye ve değişmeye başlamıştır. Bu gelişim süreci ile beraber belli başlı sorunlar da kendini göstermeye başlamıştır. İvmeli olarak yükselen insan nüfusu ve aynı ivme ile artan ulaşım ihtiyaçları, toplu yaşam alanlarında araç kazalarında artışa neden olmaktadır. Buna ek olarak trafik sorunları ve yakıt tüketimi artışı sorunları da kendini göstermektedir. Bu döngünün getirdiği sorunların yeni teknolojik kazanımların kullanımıyla çözülmesi gerektiği açıktır. Bu bağlamda, sürücüsüz araçlar veya diğer adıyla otonom araçlar konseptleri iyi bir çözüm niteliği taşımaktadır. Her çözüm kendi sorunları da beraberinde getirmektedir. Bu çözüm de beraberinde belli başlı sorunları ortaya çıkarmaktadır. Günümüzde birçok otomobil, iki aşamada incelenen dijital güvenlik yaklaşımları ile geliştirilmektedir. Bu sistemler, dış kaynaklı siber saldırılardan koruma sağlamak için gereken bir tür gömülü sistem haberleşmesi (Denetleyici Alan Ağı (CAN) gibi) kullanılarak, araç içindeki ağda oluşturulur. Bu saldırılar, kural odaklı, anomali odaklı, liste odaklı sistemler vb. gibi çeşitli yollarla tespit edilebilir. Mevcut literatür, araştırmacıların bu tür saldırıların tespiti için bazı yapay zekâ tekniklerinin kullanımına odaklandığını göstermiştir. Yapılan çalışmada CAN güvenliği için makine öğrenimi yöntemlerine dayalı bir siber saldırı tespit sistemi önerilmiştir. Sonuç olarak, karar ağacı temelli toplu öğrenme modellerinin test edilen algoritmalar içerisinde en yüksek başarımı verdiği gözlemlenmiştir.

Anahtar Kelimeler: makine öğrenimi, saldırı tespit sistemi, CAN, akıllı araç

University : **T.C. İstanbul Kültür University**
Institute : **Institute of Graduate Studies**
Department : **Computer Engineering**
Program : **Computer Engineering**
Thesis Advisor : **Prof. Dr. Özgür Koray ŞAHİNGÖZ**
Degree Awerded And Date : **MS – APRIL 2022**

ABSTRACT

MACHINE LEARNING BASED INTRUSION DETECTION FOR IN-VEHICLE NETWORKS

Parallel with the developing world, transportation technologies have started to expand and change significantly year by year. This change brings with it some inevitable problems. Increasing human population and growing transportation-needs result many accidents in urban and rural areas, and this recursively results extra traffic problems and fuel consumption. It is obvious that the issues brought by this spiral loop needed to be solved with the use of some new technological achievements. In this context, self-driving cars or automated vehicles concepts are seen as a good solution. However, this also brings some additional problems with it. Currently many cars are provided with some digital security systems, which are examined in two phases, internal and external. These systems are constructed in the car by using some type of embedded system communication (such as the Controller Area Network (CAN)) which are needed to be protected from outsider cyberattacks. These attacks can be detected by several ways such as rule based system, anomaly based systems, list based systems, etc. The current literature showed that researchers focused on the use of some artificial intelligence techniques for the detection of this type of attack. In this study, an intrusion detection system based on machine learning is proposed for the CAN security, which is the in-vehicle communication structure. As a result of the study, it has been observed that the decision tree-based ensemble learning models results the best performance in the tested models. Additionally, all models have a very good accuracy level.

Keywords: machine learning, intrusion detection system, CAN, smart car

1. GİRİŞ

Ulaşım tarih boyunca insanlığın olmazsa olmazı olmuştur. İlk çağlardan bu yana, bir noktadan diğer bir noktaya hızlı ve etkili şekilde ulaşmak büyük önem arz etmiştir. Başlangıçta kundak kullanımı ile bu ihtiyaç giderilmiş, daha sonra kabin ve aksamların eklenmesiyle otomobil kavramı gelişerek şekillenmeye başlamıştır. Günümüzde insanlık olarak akıllı ve otonom araç konseptiyle karşı karşıya kalmış durumdayız. Akıllı araç, üzerindeki sensörler ile aracın seyrüseferini, sürüş konforu ve güvenliğinin arttırılmış olduğu araç tipidir [23]. Sensör teknolojilerinin gelişmesi ile akıllı araçlar daha fazla alt sistemi takip edebilir ve bu sistemlerden toplanan veri ile karar destek sağlayabilir hale gelmiştir. Bu tanım çerçevesinde seyrüseferde sürücü kontrolü bağımsız sistemlerinde (motor arızası, sensör kontroller vs.) karşılaşılan durumlara göre karar vermesi beklenmektedir. Akıllı araçların bir üst seviyesi olarak otonom araçlar gösterilebilir. Otonom araçlar, sürücü etkisi olmadan çevrenin değişen durumuna, trafiğe ve yayalara göre sürüşünü sürdürebilen araçlar olarak ifade edilebilir [22]. Otonom araçlar akıllı arabalar kategorisinde değerlendirilmesine karşın daha hassas sensörler barındırmaktadırlar. Bu gelişmiş algılayıcı sistemi iki alt parçada incelebilir. Trafik sinyalizasyonu, hareketli nesne tespiti, haritalama ve konumlama yapıları verileri toplayan yapılardır. İkinci sistem ise karar verme veya karar destek sistemi olarak adlandırılabilir. Alınan verilerin işlenmesi ve sonuçların değerlendirilerek karar almadan oluşmaktadır. Rota planlaması, yakıt tüketimi, çevresel algılama ve trafik yönelimi gibi birçok alt sistem, karar destek yapılarının bir parçasıdır. Bu parçaların bir araya getirilip araca uygulanması ile otonom araç elde edilir.

Daha önce de belirtildiği gibi, akıllı araçların dijitalleşmesi ile araçların siber güvenliği önemli hale geldi. Siber saldırılar söz konusu olduğunda, gelen saldırı ya dış iletişim ya da iç iletişim araçlarıyla gerçekleşmektedir. Yapılan çalışmanın odağı iç iletişim ağının güvenliğinin sağlaması üzerine gerçekleştirilmiştir. Araç içerisinde standartlaşmış CAN iletişim mimarisi kendi içerisinde şifreleme veya saldırı tespiti gibi alt unsurları barındırmamaktadır. Bu eksiklik, sistemi dış müdahalelere karşı savunmasız hale getirmektedir. Önerilen çalışma ile CAN ağı üzerinde makine öğrenmesi yöntemleri ile saldırı tespiti yapılması hedeflenmiştir. Deneye tabi tutulan algoritmaların başarım sonuçları çalışmada ifade edilmiştir.

CAN, otomotiv endüstrisinde seri üretime girmiş öncü iletişim ağıdır. Standart haline gelmesi ile birlikte otomasyon teknolojilerinden biyomedikal ürünlere kadar birçok farklı alanda kullanılmaktadır. Bunun temel sebebi düşük maliyetli ve güvenilir olmasıdır. CAN, motor analizleri ve durum yönetiminden araç içi iklimlendirmeye, şarj sistemlerinden hidrolik

direksiyona kadar birçok farklı alanda kullanılmaktadır. Çalışma çerçevesinde CAN taşıyıcı mimarisi üzerinde taşınan veriler kullanılarak makine öğrenimi temelli saldırı tespit sistemi önerilmiştir.

1.1. Problem Tanımı

Problemi tanımak, onu çözmenin yarısıdır. Buradan yola çıkarak problemin olabildiğince detaylı şekilde tanımlanması ve ardından her tanım parçası için çözüm üretilmesi gerekmektedir. Temelde bir mühendislik pratiği olarak karşımıza çıkan böl ve fethet disiplini ile problem alt parçalarına, atom problem olana kadar tanımlanmalıdır.

Akıllı araçların güvenliği söz konusu olduğunda da bu yöntem izlenmiştir. Başlangıçta problem iki aşamaya ayrılmıştır. Bunlar haberleşme yöntemleri ve siber saldırı çeşitleridir. Akıllı araçlara saldırılar haberleşme vasıtaları ile gerçekleşmektedir. Bu sebeple ilk aşama haberleşme yöntemlerini alt dallara ayırarak problem çözümüne ulaşmaktır. Bu çerçevede haberleşme çeşitleri iki ana başlıkta toplanmıştır. Araç içi ve Araç dışı haberleşme.

Bu çalışma kapsamında alt problem olarak araç içi haberleşmenin saldırı tespiti ele alınmıştır. Araç içi haberleşmedeki temel unsur CAN olarak adlandırılan ağ yapısıdır. Ana hatları ile bu yapı yayın tabanlı bir iletişim sunar. Paketler şifreleme gibi güvenlik unsurlarına tabi olmadan kapalı ağ içerisinde dolaşır. Bu durumda bir saldırganın aracın iç ağına fiziksel olarak veya uzaktan erişimi durumunda sistemi saldırılara açık hale getirmektedir. Temel problem bahsedilen yapısının etkin şekilde korunmasıdır. Etkin bir sistem oluşturulması açısından çalışmada makine öğrenimi yaklaşımı önerilmektedir.

Çalışma kapsamında bahsedilen problemin bir diğer parçası ise saldırı çeşitleridir. Yapılan literatür taramasında bununla alakalı olarak içerisinde 4 farklı saldırı çeşidi barındıran veri seti görülmüş ve deneyler HCRL Car Hacking Dataset [10] üzerinde gerçekleştirilmiştir. Veri seti Güney Kore menşeili siber güvenlik laboratuvarı olan Hacking and Countermeasure Research Lab (HCRL) tarafından yayınlanmaktadır. Veri seti içerisindeki saldırı tipleri sırasıyla DOS, Bulanık Saldırı, Yanıltma Saldırısı RPM ve Yanıltma Saldırısı Gear olarak tanımlanmaktadır. Veri kümesi detaylarından ilgili bölümde bahsedilmiştir.

1.2. Literatüre Katkıları

Gelişen ulaşım ve araç teknoloji ile birlikte otonom ve akıllı araç konseptleri de günlük yaşantının içine sirayet etmektedir. Bu durum, otonom ve yapay zeka destekli sürüş araştırma alanına olan akademik ve endüstriyel ilgiyi de arttırmaktadır. Yapılan çalışma da bu araştırma

alanının bir alt dalı olarak tanımlanabilir. Araçların dijitalleşmesi ve akıllı hal gelmesi haberleşme teknolojiler ile mümkün olmaktır. Özellikle araç içi sensörlerin kendi aralarında iletişimin sağlaması kilit bir unsur olarak değerlendirilebilir. Bu çerçevede yapılan çalışma ile birlikte bu iç ağın güvenliğinde Makine Öğrenimi mekanizmaları deneyimlenmiş sonuçları paylaşılmıştır. Buradan yola çıkarak bu sonuçlar ilerleyen safhalarda endüstri ve akademi için bir katkı oluşturmaktır.

Katkı 1. Yapılan çalışmada uygulanan yöntemler genel isterler açısından uygunluk göstermektedir. Burada bahsi geçen isterler paketlerin denetlenmesi safhasının hızlı ve yüksek doğruluk ile saptanmasıdır. Bu doğrultuda seçilen makine öğrenimi algoritmaları, uygulanabilirlik safhasında bir kıyas tablosu ortaya koymaktır.

Katkı 2. Makine öğrenimi ve Çok Katmanlı Algılayıcı yapılarının belirlenmiş olan veri ve problem uzayı üzerindeki etkileri incelenmiş ve deney sonuçları ortaya konuştur. Bununla birlikte bahsedilen ve Çok Katmanlı Algılayıcı yöntemlerinde sonuçları değerlendirilmektedir.

Katkı 3. Her iki yönteminde tek bir çatı altında denenmesi başlı başına bir önem arz etmektedir. Bu önem kıyas yapabilme imkanından doğmaktadır. Yapılan çalışma ile birlikte uygulanan tüm yöntemler kıyaslı olarak incelenmektedir.

Katkı 4. CAN Taşıyıcı yapısının detaylı incelenmesi ve veri setinin ön işlemeye tabi tutulması ve içerisinde öz niteliklerin incelenmesi çalışma alanı için genel bir görü kazanılması ve olası diğer akademik ve endüstriyel çalışmalara katkı sağlaması bağlamında önemi arz etmektedir.

Katkı 5. Çalışmanın bir diğer önem arz eden katkısı da endüstri açısından değerlendirilebilir. Burada saldırı tespiti gerek akıllı ve otonom araçların sürüş güvenliğini direkt olarak etkilemekte gerekse araç içi verilen ve sensörlerine güvenliği için önem arz etmektedir. Yapılan çalışma direkt olarak bahsedilen durumlara katkı sağlamaktadır.

1.3. Tezin Organizasyonu

Tez çalışması beş bölümden oluşmaktadır. Birinci bölümde, problem tanımı yapılmıştır. Yapılan çalışma tanıtılmıştır, amacı ve önemi anlatılmıştır ve literatüre katkısından söz edilmiştir. İkinci bölümde, tez çalışmasının ana kaynağı olan Makine Öğrenmesi algoritmaları, Çok Katmanlı Algılayıcı ve kullanılan veri setinden bahsedilmiştir. Üçüncü bölümde, çalışmada kullanılan yöntemler, çalışmada kullanılan teoriler, yaklaşımlardan ve bunların nasıl uygulandığından, amaçlarından bahsedilmiştir. Dördüncü bölümde, yapılan çalışma önerilen yöntemle ilgili detaylardan, bileşenlerden, yöntemin akış diyagramından detaylı bir şekilde bahsedilmiştir. Son bölümde ise yapılan testler ve sonuçları detaylı bir şekilde belirtilmiştir.

2. ÖN BİLGİLER ve LİTERATÜR ARAŞTIRMASI

Bu bölümde araştırması yapılan çalışmanın temel tanımları ve çalışma için yapılan literatür taraması sunulmuştur.

2.1. Akıllı Araçlarda Saldırı Tespiti

2.1.1. Araçlar ve Ulaşım

Ulaşım insanlık için kritik bir gerekliliktir. İnsanlık ile beraber bir noktadan başka bir noktaya kolay, etkili olarak ulaşım sağlamak her daim önem arz etmiştir. Başlarda doğanın kendisinden, yani binek hayvanlarından yararlanılarak giderilen bu ihtiyaç, ilerleyen zamanlarda bu bineklere bir takım kabin ve araç-gereç eklemeleri ile gelişmiş ve yavaş yavaş, araba kavramı oluşmaya başlamıştır. Araba temelde 4 teker üzerinde bir güç üretecinin etkisi ile hareket eden içerisinde bir grup insanın taşınmasını sağlayan bir ulaşım aracıdır. Modern otomobilin hızı kullanım kabiliyeti ve bakım masrafları ve iaşesi bir binek hayvana göre daha sürdürülebilir olmasından dolayı kullanımı kronolojik olarak giderek artmaya başlamıştır. Seri üretime geçen ilk araç, olan Henry Ford'un geliştirdiği Ford Model T1 ile birlikte standart haline gelmeye başlayan modern araba, evrim sürecine devam ederek günümüzdeki halini almaya başladı. Günümüzde ise akıllı araç ve otonom araç, konsepti ile farklı bir bakış açısı kazanmaktadır.

Araçlardaki gelişmeler transistör teknolojisi ile birleşmeye başladıkça araçlarda da dijital devrim başlamıştır. Bununla birlikte araçlara elektronik parçalar girmeye başlamıştır. Her elektronik parça gibi bu parçalar arasında da haberleşme ve haberleşme ile birlikte güvenlik ihtiyacı hasıl olmuştur. Araçlardaki karmaşıklık arttıkça bu ihtiyaç, sistemleri de gelişmeye zorlamaya başlamıştır. Araç içi haberleşmede CAN gibi standartlar ortaya çıkmıştır. Bu da araç içi parçaların haberleşmesini kolaylaştırmış ve global ölçekte üretim standartı haline gelmiştir. Fakat sistem kendi içerisinde güvenlik unsuru barındırmayan bir haberleşme yöntemidir. Temelde bunun sebebi haberleşmenin hızlı şekilde gerçekleşmesi gerekliliğidir. Bu güvenlik eksikliği de sistem içinde açıklara sebebiyet vermektedir.

Akıllı Araç: Akıllı araçlar üzerinde sensörler barındıran ve sensörler vasıtasıyla ürettiği büyük veriyi kullanarak verimliliği, sürüş-kullanım kolaylığını hedefleyen araçlardır [23]. İlerleyen sensör teknolojileri ile beraber akıllı araçların veri yönetimi de gelişmektedir. Bununla birlikte araç kontrolleri dijitalleşmeye başlamış ve sürücü etmen kullanımdan otonomi etmen kullanıma bir geçiş sağlanmıştır. Bu tanım çerçevesinde aracın kullanıcı bağımsız sistemlerde karşılaşılan duruma göre karar vermesi beklenmektedir.

Otonom Araç: Sürücüden bağımsız olarak çevre şartlarına, araç trafiğine ve yaya geçitlerine göre

sürüşünü gerçekleştirebilen ve sürdürebilen araçlar olarak tanımlanabilirler [22]. Otonom araçlar akademik ve endüstriyel çevrelerin 1980'lerden başlayarak ilgi odağı olmuş bir araştırma konusudur. Özellikle Defense Advanced Research Projects Agency (DARPA)'nın [1] Grand Challenge çerçevesinde düzenlediği otonom araçlar yarışmaları, konuya ilgiyi artırmıştır. Gelişen sensör teknolojisi ve Yapay Zeka algoritmaları ile günümüzde önemi artan ivmeyle devam etmektedir. Sensörler vasıtasıyla çevrelerindeki durumların verisini toplayıp ve ona göre sürüş sağlamaktadırlar. Temelde bu sistem iki farklı yapıdan oluşur. Algılayıcı yapısı çevreden sensörlerle alınan veriyi ve bu verinin miktarını ve çeşitliliğini oluşturur. Algılayıcılar kendi içerisinde trafik sinyalizasyonu, hareketli obje tespiti, haritalayıcı ve konumlayıcı yapıları veriyi toplayan yapılardır. İkinci yapı ise karar destek sistemi olarak ifade edilebilir. Bu yapı alınan verinin işlenmesi ve bu doğrultuda tüm sistemlerin karar oluşturulmasından sorumlu olan yapı olarak tanımlanabilir. Aracın rota planlanması, yakıt tüketimi, çevre algısı, trafik yönelimi gibi birçok alt sistem karar destek yapısının bir parçasıdır. Bu parçaların birleşmesi ve bir arabanın üzerinde uygulanmasıyla otonom araç yapısı elde edilmiş olur.

Yarı Otonom Araçlar: Otonom araçlardan sonra ortaya çıkmış ve insan-araba etkileşimi ile insanın araba sürüş sorumlulukları azaltmayı aynı zamanda otonom araçların tam otonomisini kısıtlayarak, sürücü kontrolünde ve otonomi destekli bir yapı ortaya koymuştur. Bu sayede insan-araba arasında etkileşim ile sürücünün daha etkili araç kullanımı sağlanır ve sürüş güvenliği artırılmış olur. Bu konu çerçevesinde L. Friedman üç temel fikirden yola çıkarak yarı otonom araç sistemini önermiştir [2]. Bu üç fikir:

- *Sürüş kolay bir iştir. (Driving task is easy)*
- *İnsanlar sürüşte iyi değillerdir. (Humans are not good at driving)*
- *İnsan ve otomasyon beraber iyi çalışamazlar. (Humans and automation don't mix well.)*

Önerilen sistem sürücüyü takip ederek gerekli durumlarda müdahale sürücüye uyarılar vererek veya bizzat sürüşe etkide bulunarak bahsedilen iyileştirmeleri sağlamış olur. Ek olarak birçok yolda sürüş tam otonom sistem tarafından gerçekleştirilebilmektedir ve yer yer sürüş sürücüye devredilmektedir. Sürüş sırasında sürücünün kafa hareketleri, göz hareketleri takip edilerek sürücünün yola odaklanması da ölçülmektedir. Bu sayede sürücünün aracı sürerken sürüş dışında gerçekleştirdiği etkileşimler kayıt altına alınmakta ve uyarılar yapılmaktadır. Ayrıca çevresel faktörler değerlendirilmektedir. Yolun kenarında bekleyen bir yayanın konumu, sürücünün ilgi odağı gibi faktörler göz önüne alarak risk hesaplaması gerçekleşmektedir.

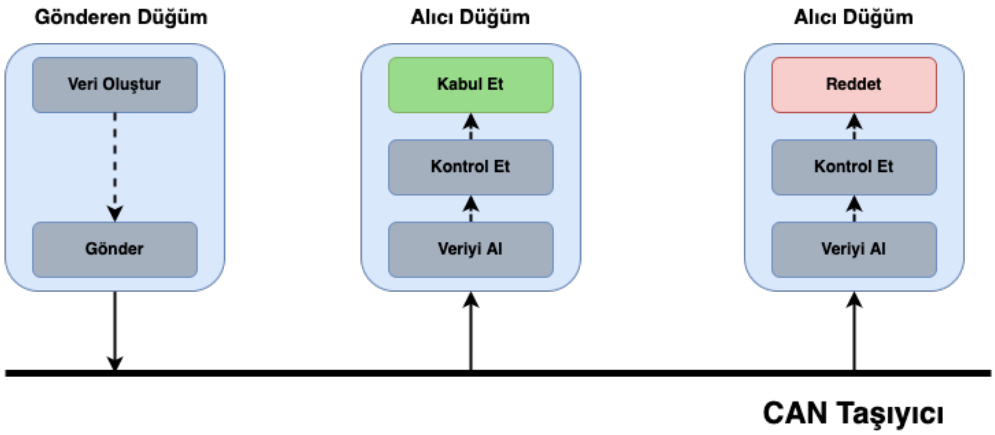
2.1.2. Haberleşme

Akıllı araçlarda, CAN yapısı ile sensör ve aktüatör arasında iletişim kurulması gereken birçok

cihaz bulunmaktadır. İletişim mimarisi için farklı yapılar sağlanarak sensör ve aktüatörlerin haberleşmesi sağlanmış olur. Örnek olarak, olay tabanlı iletişim modellerinde, sensörler veya aktüatörler tetikleyici bir olay meydana geldiğinde iletişim ağı üzerinden mesaj gönderebilir [3] [4]. Öte yandan, yayın aboneliği ve açık yayımlama iletişim paradigmasında bu cihazlar farklı tür içerikleri ağ üzerinden takip edebilmek için kayıt olabilirler. İletim mekanizması da bunları [5] [6] 'de belirtildiği gibi abonelik kriterlerine göre iletilir. Araç iletişimi iki alt grup altında incelebilir. Bunlar araç dışı ve araç içi iletişim sistemleridir. Araç içi iletişim sistemleri, sensörler ve CAN taşıyıcı ağı üzerinden karar birimleri arasındaki iletişimi ve bilgi alışverişini ifade eder. Çalışmada önerilen IDS için araç içi iletişim sisteminin güvenliğinin analizi hedeflenmektedir. Diğer bir iletişim türü, araçlar arası iletişim sistemleridir. Bu çok yönlü iletişim sisteminde araçlar, örgü tipi ağlar aracılığıyla diğer araçlara, trafik ışıkları gibi altyapılara, yayalara ve bilgi işlem merkezlerine bilgi paylaşır.

2.1.2.1. Araç İçi Haberleşme

Bahsedilen araç içi haberleşme yapısı için, CAN taşıyıcı ağı ekseriyetle Şekil 2.1'de gösterildiği gibidir. Bileşenlerin sayısı artarsa, fiziksel kabloların kullanımı ek sorunlara sebebiyet verebilir, bu nedenle araç içi iletişimde kablosuz haberleşme sistemleri de tercih edilebilmektedir.



Şekil 2.1- CAN Taşıyıcı Yapısı

Ultra Geniş Bant, radyo dalgalarını kullanan bir kablosuz iletişim protokolüdür. UGB, veri iletimi için geniş frekans aralığı kullanmaktadır. Bunu gerçekleştirirken, geleneksel yöntemlere kıyasla daha fazla veri iletebilir. Fakat UGB veriyi yüksek frekans ve geniş frekans aralığı sebebiyle kısa mesafede iletebilmektedir. UGB, küçük alanlı ortamlarda kablolu sistemler kadar hızlı ve sağlıklı veri iletimi gerçekleştirebilmektedir.

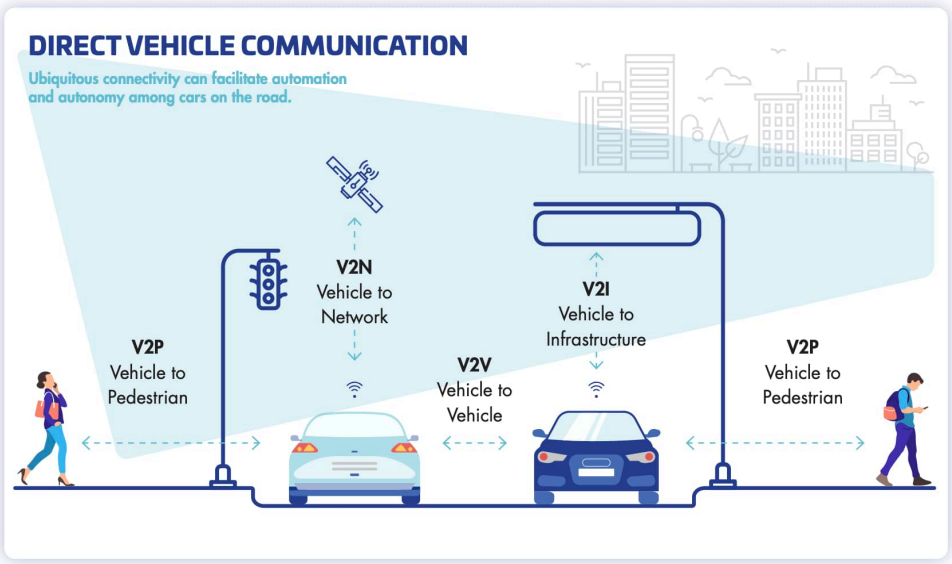
Araç içerisindeki sensörler ile bunları kullanarak karar desteği ve araç içi kontrol sağlayan ECU'lar arasındaki bağlantı bu yapı ile sağlanmaktadır. CAN taşıyıcı yapısı, ağ paketlerinde 8

baytlık veri paketleri taşımaktadır. Sensörlerin konumu, özellikle kablosuz sensör ağları konseptlerinin kullanımında önem arz etmekte ve sistem geliştirilmesinde etmen olmaktadır [7]. Belli başlı araştırmalarda, konum doğrudan sensörlerin kapsama alanına bağlı olarak değişmektedir. Bununla birlikte, bir araçta, konum doğrudan taşıyıcı arabirimin yoluna bağlıdır.

Ağ protokolü yayın mimarisi her birimden verileri alır ve tüm birimlere gönderilir. Verileri ilgili olan birimler kabul eder ve işleyerek kullanır. Araç içindeki iletişimin anlık olması gerekliliğinden, veri paketlerinde şifreleme sistemleri kullanılmamaktadır. Bu durum, herhangi bir kötü niyetli ara unsurun saldırmak için dahili ağa erişmesine ve manipülasyon gerçekleştirmesine olanak verir. Bu durum sistemin içeriden korunması gerekliliğini ortaya koyar. Bu çalışmada, makine öğrenme algoritmaları problem için bir saldırı tespit sistemi olarak önerilmektedir.

2.1.2.2. Araç Dışı Haberleşme

Şekil 2.2'de görüleceği üzere Araç Dışı Haberleşme aracın kendi bünyesindeki parçalar dışındaki dünya ile haberleşmesidir. Bu çerçevede RSU adaptörleri özellikle alt yapı iletişiminde önem arz etmektedir.



Şekil 2.2-Araç Dışı Haberleşme Yöntemleri [8]

Vehicle-to-Vehicle (V2V): Araçtan araca (V2V) haberleşme, araçların hızları, konumları ve rotaları hakkında kablosuz olarak bilgi alışverişinde bulunmalarını sağlar. V2V iletişimin kullandığı teknoloji, araçların tüm yönlü mesajları yayınlamasına ve almasına olanak sağlayarak, çevredeki diğer araçların tüm yönelimli farkındalık oluşturmasını sağlar. Uygun yazılımlar ile desteklenmiş araçlar potansiyel kaza durumlarını tespit etmek için etraftaki araçlardan gelen verileri kullanabilmektedir. Sonrasında bahsedilen yapı sürücüyü uyarmak için görsel, haptik

ve/veya sesli uyarıları kullanarak bildirim gerçekleştirir. Bu tip sistemler sürücünün olası kaza durumlarından kaçınmasına yardımcı olmaktadır [3].

Vehicle-to-Infrastructure (V2I): Akıllı Ulaşım Sistemlerinin bir parçası olan Araç'tan Altyapıya, yol ve hava şartları gibi bilgilerin kablosuz bağlantılar aracılığıyla araca iletilmesini ifade etmektedir. Bu teknoloji vasıtasıyla uygun park alanları, yakıt ücret bilgileri, çevredeki şarj istasyonları, geçiş¸ önceliği olan araçların geçişi için bilgilendirme, kaza bilgisi, yolun iklim şartlarına göre durumu gibi çeşitli bilgiler iletilmektedir. Amerika Birleşik Devletleri Ulaştırma Bakanlığı tarafından paylaşılan rapora göre mevcut trafik kazaların %80'ini bu sistem sayesinde önlenebilir olmaktadır. Buna ek olarak seyahat sırasında geçirilen zamanın %10'u trafikte harcanırken, trafiğin %12'si ise park arama sırasında oluştuğu saptanmıştır. Bunlara ek olarak şehir içindeki yakıt tüketiminin %17'si ise trafik ışıklarında beklerken gerçekleştiği raporda belirtilmektedir [4].

Vehicle-to-Pedestrian (V2P): Yaya ve araç iletişimidir. Bu çerçevede bir akıllı telefon araştırması yapılmıştır [19]. Akıllı telefonların yol bağlamında kullanılması VRU'ları hızla arttırmaktadır. Akıllı telefon kullanımının etkisiyle ilgili riskleri azaltmak Trafiğin göz önünde bulundurulması gereken bir durumda, bir çarpışma tahmin algoritması kullanılması bu iletişim çerçevesinde değerlendirilebilir.

Vehicle-to-User (V2U): Araçtan sürücüye/kullanıcıya¸ (V2U), sürücü ve araç dahil yolcuları arasındaki iletişimi ifade eder. Bu iletişim bilgi, eğlence ve seyahat deneyiminin kişiselleştirilmesi ile alakalıdır. İnsanların dijital deneyimini evden ve işyerinden arabaya genişleten bu teknoloji üzerinde hali hazırda birçok araç üreticisi araştırma ve geliştirmeye yatırım yapmaktadır [20]. Temelde araç içi medya sistemi veya araç içi eğlence sistemleri ile ifade edilir.

Vehicle-to-Network (V2N): Araçla belki birden fazla frekansta, araca erişim sağlayan şebeke operatörü arasındaki bağlantıyı temsil eder. Bu iletişim kritik olmayan ve yolcuya dayalı deneyimleri iyileştiren ve aynı zamanda internet bağlantısı için bir erişim noktası olarak hizmet veren bulut tabanlı hizmetler bütünü olarak ifade edilebilir.

2.2. Makine Öğrenimi

Makine öğrenimi veriden yola çıkarak, her olası durumu ve sonucu el yordamı ile programlamadan, istatistik ve matematik yöntem ve teknikleri ile, verideki örüntü yapılarının

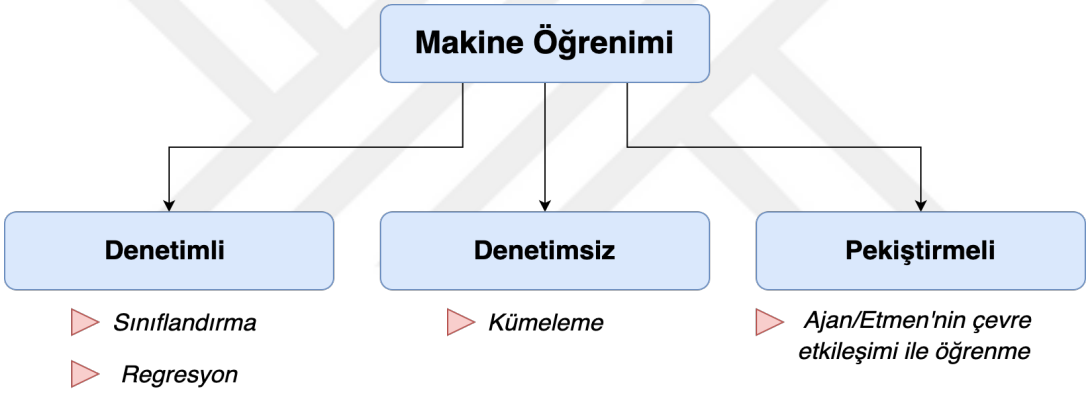
ortaya çıkarılması ve bununla bağlantılı olarak çeşitli problemlere çözüm üretilmesini hedefleyen bir Yapay Zeka alt araştırma dalıdır.

Makine öğrenimi en temelde örüntü keşfi için kullanılan yöntemler ve teknikler bütünü olsa da kendi içerisinde çeşitli problemlere spesifik olarak geliştirilen algoritmalar barındırmaktadır. Bahsedilen problemler ana başlıkları ile Regresyon, Sınıflandırma, Kümeleme ve Ajan/Etken optimizasyonu olarak tanımlanabilir.

Bahsedilen problemler ve uygulanan çözümler alt başlıklarda detaylı olarak incelenmiştir.

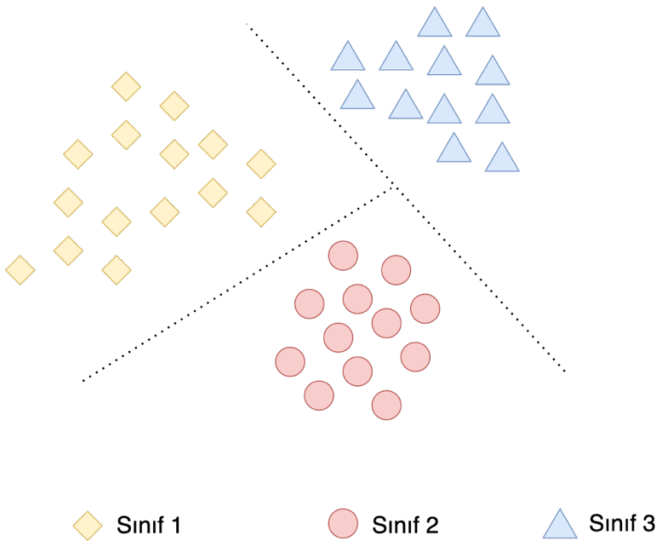
2.2.1. Makine Öğrenimi Yöntemleri

Makine öğrenimi algoritmalarını kendi içerisinde üç temel başlıkta incelenmektedir. Bunlar sırası ile Denetimli Öğrenme, Denetimsiz Öğrenme ve Pekiştirmeli Öğrenmedir. Tem alt problemleri Şekil 2.3'te gösterilmektedir.



Şekil 2.3- Makine Öğrenimi Teknikleri

Denetimli Öğrenme: Makine Öğrenimi bahsedildiği üzere açıkça programlama yapılmadan eldeki verileri kullanarak belirli problemlere esnek çözümler üretmeyi hedeflemektedir. Bu çerçevede Denetimli öğrenme yöntemleri eldeki verinin insan denetiminden geçmiş olduğu ifade etmektedir. Literatürdeki tabiri ile etiketleme olarak adlandırılan bu işlem ile, eldeki veri onu temsil eden veya o veriden yola çıkarılarak elde edilen sonuçların insan tarafından nitelendirilmesi ve tahmini veya örüntü çıkarımı yapılacak öz niteliğin belirlenmesi ile ifade edilebilir. Diğer bir deyişle, Şekil 2.4'te ifade de edildiği gibi, sınıflar arasındaki sınırların belirlenmesidir. Örneğin bir evin değerinin tahmin edilmesi, bunu yaparken evin özelliklerinin kullanılması gösterilebilir. Özetle eldeki verilerin etiketlenmiş olduğu durumlarda uygulanan Makine Öğrenimi Yöntemini bu başlık altında incelenmektedir. Bahsedilen yöntemlere örnek olarak, KNN, Karar Araçları, Naif Bayes, Lineer Regresyon ve daha niceleri sayılabilir.

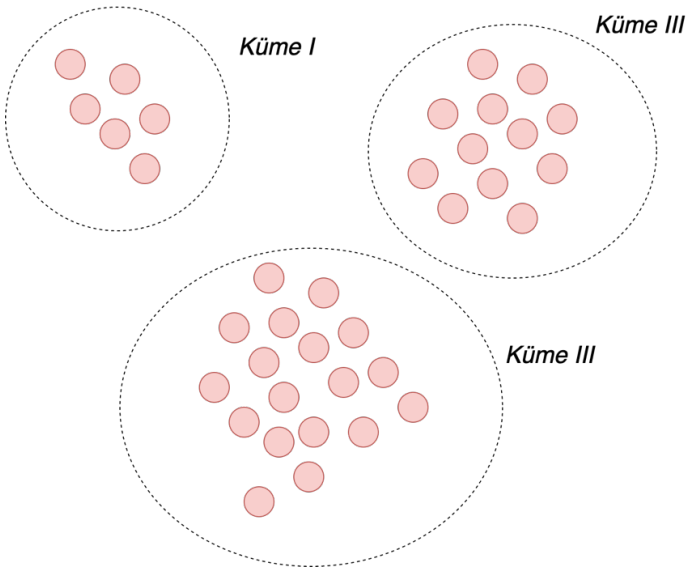


Şekil 2.4-Denetimli Öğrenme Yapısı

Denetimsiz Öğrenme: Eldeki verinin insan denetiminden geçmemiş, literatürdeki ifadesi ile etiketlenmemiş olma durumunu ifade eden Makine Öğrenimi yöntemleri bütünüdür. Denetimsiz öğrenme yöntemleri kendi içerisinde çeşitli alt gruplara sahiptir. Bunlar başta Şekil 2.5'te görüleceği üzere Kümeleme yöntemleri olmak üzere, İlişkilendirme ve Boyut Düşürme olarak sıralanabilir. Denetimsiz öğrenme çerçevesinde en çok kullanılan yöntemler kümeleme yöntemleridir. Çeşitli Yöntemler olmasına karşın en çok K-Means ve K-Medoids algoritması kullanılmaktadır.

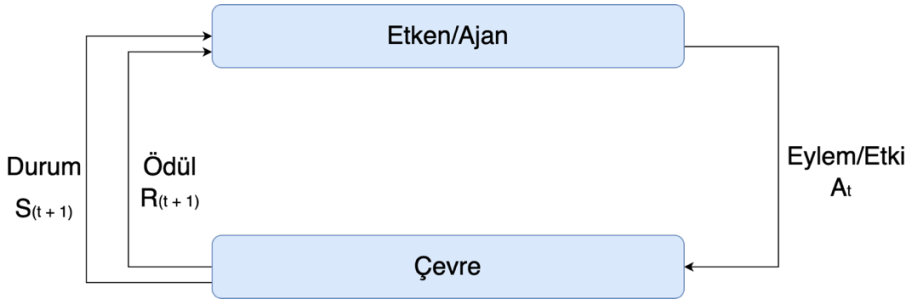
İlişkilendirme yöntemleri ise özellik market sepet analizlerinde kullanılan yöntemlerdir. Bu yöntemler kullanıcı alım eğilimlerinin analizlerinde kullanılır. Burada da FP-Tree ve Apriori Algoritmaları genel geçer olarak kullanılmaktadır.

Son olarak Boyut düşürme algoritmalarında verini boyutu istatistiksel yöntemler ile azaltılarak verini en uygun temsili elde edilmeye çalışır. Buradaki amaç hafızadan kazanç veya veri görselleştirme olabilir. Bu çerçevede en bilinen yöntem PCA olarak ifade edilir.



Şekil 2.5-Denetimsiz Öğrenme Yapısı

Pekiştirmeli Öğrenme: Göreli olarak uygulama alanı diğer iki yönteme göre daha azdır. Buradaki temel yaklaşım ajanın-etmenin çevrenin etkilerini karşı tepki vermesi ve bunun sonucunda ödül ceza yaklaşımları ile ajanın çevreyi öğrenmesi olarak değerlendirilebilir. Bu yöntemin genel hatları Şekil 2.6'da ifade edilmiştir.



Şekil 2.6-Pekiştirmeli Öğrenme Yapısı

2.3. Literatür Araştırması

Bu bölümde tez çalışmasında yapılan literatür araştırmasına ve ilgili çalışmalara değinilmiştir.

CAN taşıyıcı yapısı etkin bir yapıdır fakat güvenlik açısından da eksik bir iletişim protokolüdür. Bu eksiklik, saldırı-sızma tespit sistemleri ile kapatılmalıdır. Ek olarak, araçlara üzerinde uygulanabilen saldırıların sayısı da kısıtlıdır. Bilinen saldırıların sayısının kısıtlı olması, henüz tespit edilmemiş nice saldırı olabileceğini de ortaya koymaktadır. Yapılan çalışmada Rekabetçi Üretici Ağlar Derin Öğrenme mimarisi ile geliştirilmiş bir saldırı tespit sistemi

önermektedir [10]. Bu yapı, potansiyel olarak bilinmeyen saldırılara karşı da çözüm üretebilir. GAN'lar yapıları gereği üretici ve ayrıştırıcı olmak üzere iki kısımdan oluşmaktadır [11]. Bu yapı ile modelde de benzerlikler tespit edilebilmektedir. Bu iki farklı ağ yapısı, özellikle görüntü verileri üzerinde yapılan çalışmalarda etkili sonuçlar vermektedir. Çalışmada CAN taşıyıcısı üzerinde elde edilen onaltılık (hexadecimal) verilere One Hot Encoding yöntemi uygulanmıştır. Rekabetçi Üretici Ağlar mimarisi ile bilinmeyen saldırılara karşı çok boyutlu matris formatındaki veriler alınmaya çalışılmıştır.

Grafik işleme teknikleri, yüksek maliyetli hesaplamaların daha etkin olarak gerçekleştirilmesine olanak sağlamaktadır. Evrişimli Sinir Ağları hem grafik ölçekli hesaplama hem de geri yayılım algoritması nedeniyle yüksek hesaplama maliyetlerine sahiptir. Buna karşın karmaşıklığı yüksek olan problemlere karşı yüksek başarım göstermektedir. Çalışmada, Evrişimli Sinir Ağları, siber saldırılara karşı araçların IDS'i olarak test edilmiştir [12]. Mimari, Evrişimli Sinir Ağı, LSTM ve Azaltılmış Başlangıç ile ResNet mimarilerini kullanmıştır. Ayrıca Standart Makine Öğrenmesi yapıları denenmiştir. Makine öğrenmesi algoritmaları arasında SVM, K En Yakın Komşu, NB ve Karar Ağaçları sınanmıştır. Bunlar arasında en karmaşık yapıda çalışan Azaltılmış Başlangıç ile ResNet, en yüksek başarımı göstermiştir.

Çalışmada CAN taşıyıcı yapısı üzerinde çalışan bir IDS geliştirilmiştir [15]. CAN veri yolundaki veriler, kimlik doğrulaması yapılmadan bir noktadan diğerine gider. Bu da sistemindeki güvenlik açıklarını ortaya çıkarır. Çalışma Fuzzy ve Dos saldırıları üzerinde gerçekleştirilmiştir. Dos saldırılarında CAN ID verisine 0x000 değerinin hâkim olduğu görülmüştür. Bulanık saldırı (Tuesday) çeşidinde, CAN Kimliği ve verilerde rasgelelik gözlemlenmiştir. Bu saldırıyı belirlemek için K En Yakın Komşu ve SVM algoritmaları denenmiş ve K En Yakın Komşu algoritmasının daha başarılı olduğu tespit edilmiştir.

CAN taşıyıcı yapısı üzerindeki iletişim ve yayınlaması hızlı fazda gerçekleşir. Genel olarak, birimler birbirleriyle belirli bir desende iletişim kurarlar. İncelenen çalışmada bu deseni algılamak ve içinde anomali olabilecek paketleri belirlemek için bir IDS yapısı önerilmektedir. CAN taşıyıcı ağı üzerindeki akan veri izlenmiştir [16]. Bunun sonucunda belirli CAN ID'lerin belirli zamanlarda belirli mesajlar gönderdiği tespit edilmiştir. IDS sistemi iki katmanlı olarak ifade edilmiştir. Eğitim safhasında paket iletimine bağlı olarak, CAN ID'ler arasındaki olası tüm işlemler bir matris ile gösterilmiştir. Ardından test aşamasından bu matris ile karşılaştırılarak eşleşmeyen mesajlar ve ID'ler belirlenmiştir.

Derin Öğrenme yöntemlerindeki ilerlemeler ile konu özelindeki mimariler de özelleşmeye başlamıştır. Derin öğrenme yapılarındaki yüksek veri isteri ve karmaşıklığı olan problemleri çözmedeki başarısı nedeniyle yapılan çalışmada denetimsiz derin öğrenme mimarisi oluşturulmuş

ve CAN taşıyıcı yapısındaki anomaliler tespit edilmeye çalışılmıştır [19]. CANet olarak isimlendirilen mimari iki farklı yapay sinir ağı içermektedir. Bunlardan ilki LSTM'dir [5]. LSTM, ardışık veriler arasındaki desenlerin keşfedilmesini sağlayan bir yapıya sahiptir. Bu sayede belli bir zaman-bölüm dilimindeki tekrarlamalardan öğrenebilir. Diğer yapı ise Otomatik Kodlayıcı (Autoencoder) [6] mimarisidir. Gelen verileri kodlayıcı ile özütlenir veya diğer bir ifade ile vektör haline getirilir. Modelin tahminleme mekanizmasında, ağdaki veri paketinde bir anormallik olup olmadığını belirlemek için veri modele verilir ve çıkışta alınan verilerle sapma değeri hesaplanır. Hata değeri belirli bir eşiği aştığında, bu durum anomaliyi ifade etmektedir. Yapılan çalışma ile beş farklı saldırı tipi test edilmiştir.

CAN taşıyıcı güvenliği konusunda birçok akademik değerlendirme hali hazırda mevcut bulunmaktadır. Çalışmada CAN taşıyıcı yapısının güvenliğinin IDS sistemleri vasıtasıyla sağlanması analiz edilmiştir [20]. Genel olarak ECU'ların durumu araç üzerinde ifade edilmiş ve örneklendirilmiştir. Ardından mimari anlatılmış ve sistemdeki zafiyetlerden izah edilmiştir. Son olarak Anomali Tabanlı ve İmza Tabanlı Çözümler incelenmiştir.

2.4. Veri Kümesi

Veri, ham bilgi olarak adlandırılan, aktör(agent) ve çevre arasındaki etkileşimden doğan yapıdır. Ham olarak adlandırılan veri işlenmediği yahut anlamlandırılmadığı takdirde bir önem arz etmez. İlerleyen Yapay Zeka teknolojileri ile birlikte veri geliştirilen algoritmaların iyileşmesinde itici gücü oluşturmaktadır. Ek olarak bu algoritmalar sayesinde veri işlenerek son hali olan bilgi formunu alabilmektedir. Buradaki temel çıkarım Yapay Zeka ile birlikte verinin önemi önemli ölçüde artmış ve kullanım şeklinde göreli olarak bir eksen kayması olmuştur.

Derin Öğrenme mimarileri için verinin niteliği ve niceliği büyük önem taşımaktadır. Özellikle verinin niceliği, diğer bir ifade ile miktarı kritik öneme sahiptir. Yapılan çalışma çerçevesinde göreli olarak büyük veri kullanılmıştır.

2.4.1. Yararlanılan Veri Kümesi

Toplamda yaklaşık 16,5 milyon örneklem incelenmiştir. Verilerin kaynağında normal mesaj olarak etiketlenmiş fazladan 1 milyon örnek vardır. Çalışmada saldırı ifade etmeyen mesaj örneği sayısı yeterli görüldüğü için eklenmemiştir [21]. Ayrı veri kümeleri şeklinde dört farklı saldırı tipi mevcuttur. Bu veri kümelerindeki örneklem dağılımı homojen olmadığı için, kullanılan modeller için veriler farklı şekillerde kullanılmış ve test edilmiştir. İlk aşamada veri setleri tek bir veri seti altında birleştirilmiş ve belirlenen modeller bu veriler ile eğitilmiştir. Ardından her saldırı türü için homojen veri kümeleri oluşturulmuştur. Bu homojen veri kümeleri, saldırı olarak etiketlenen

mesaj sayısı kadar normal olarak etiketlenen mesajların bir araya getirilmesiyle sağlanmaktadır. Örnekleme prosedürü için normal mesajlar rastgele seçilmiştir. Bu işlemler sonucunda modeller için tüm örneklerin ve her saldırı için ayrı, homojen veri kümelerinin bulunduğu veri kümeleri oluşturulmuştur.

2.4.2. Yararlanılan Veri Kümesi Öznitelikleri

Verilerin niceliği Tablo 2.1'de ifade edilmiştir. Her saldırı türlerünün ayrı özellikleri vardır. DOS saldırısı için mesajlar her 0.3 milisaniyede bir CAN ID'si '0000' olan düğümden yapıldığı görülmektedir. CAN ID'nin, modellerin başarısını önemli ölçüde arttırabileceği ve ayrılmış veri kümesinde bir yanılsama oluşturabileceği durumdan dolayı, CAN ID öz niteliği tüm veri kümelerinden kaldırılmıştır. Ayrıca, Yanıltma saldırılarında yalnızca belirli CAN ID'ler saldırı yaptığı görülmüştür. Yine bu sebepte yanılsamaya sebebiyet verebileceği için belirletilen öz nitelik veriden çıkarılmıştır. Bu saldırı için de her milisaniyede bir saldırı paketi gönderilir. Bulanık saldırı durumda ise saldırının içeriği ve CAN ID'leri de rastgele olarak ortaya çıkmaktadır ve her 0,5 milisaniyede bir mesaj gönderilmiştir. Tüm bu saldırılar OBD portu üzerinden simüle edilerek saldırıların verileri elde edilmiştir.

Tablo 2.1- Veri Kümesi Örneklem Miktarı

Saldırı / Veri Miktarı	**Saldırı Verisi**	**Normal Veri**
Yanıltma Saldırısı RPM	654.897	3.966.805
Yanıltma Saldırısı Gear	491.847	3.347.013
Bulanık Saldırı	597.252	3.845.890
DOS	587.521	3.078.250
Toplam	2.331.517	14.237.958

3. YÖNTEM

Akıllı araçlarda saldırı tespiti çeşitli şekillerde gerçekleştirilebilmektedir. Temelde iki farklı yaklaşım ile, belirlenen problem uzayına çözüm önerilmektedir. Bunlardan ilki kural tabanlı sistemlerdir. Kural tabanlı sistemler, hız ve anlaşılırlık bakımından ikinci olarak bahsedilecek istatistik ve Makine Öğrenmesi yöntemlerinin göre etkindir fakat konu başarıma geldiğinde Kural tabanlı sistemlerin kesinliğinin yüksek olması, belli başlı olağandışı durumlara veya değişen durum ve veriye ayak uydurmasını neredeyse imkânsız hale getirmektedir. Bu sistemler belirlenen kuralların olabilme olasılıklarına göre bir ağaç gibi dallanarak sonuç vermektedirler. Bu tip sistemler sıklıkla el ile tekrar kural tanımlanmasına ihtiyaç duymaktadır.

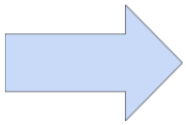
Yapılan çalışmanın kapsamı ikinci olarak bahsedilen Makine Öğrenmesine dayalı sistemlere girmektedir. Makine Öğrenmesi elde bulunan verilerden bir sonuç çıkarılarak bu sonucun kullanılmasıyla tahmin, sınıflandırma, tavsiye veya gruplama gibi işlemlerin göreli olarak yüksek doğruluk ile yapılmasıdır. Çalışma çerçevesinde "Hacking and Countermeasure Research Lab" kurumunun akıllı araç hack veri kümesi kullanılmıştır. Bu verisindeki saldırı tiplerinin yüksek doğruluk ile sınıflandırılması üzerine bir sistem önerilmektedir. Bu çerçevede makine öğrenmesi yöntemleri denemiş ve sonuçları ortaya konulmuştur. Veri uygun formata getirilmiştir.

Ardından veri içerisinde değer numerik hale getirilmiştir. Burada iki farklı yöntem izlenmiştir. İlki standart Discrete Encoding olarak adlandırılan bir nominal değeri 0'dan farklı olan değer sayısı kadar tam sayı atanarak yapılan değeri sayısallaştırma yöntemidir. Diğeri ise "One Hot Encoding" olarak adlandırılan ve her farklı değer bir öz nitelikmiş gibi düşünülerek var olma durumuna göre değerin 0 veya 1 olduğu durumdur.

3.1. One Hot Encoding

One Hot Encoding eldeki sayısal olmayan verilerin makinenin model için kullanılabilecek hale gelmesini sağlayan bir çevirme yöntemidir. Buradaki temel amaç sayısal olmayan verinin sayısal uzaya en doğru şekilde taşınmasıdır. Bunun için sayısal olmayan veriler ayrı bir öznitelik gibi düşünülerek eldeki veri kümesine eklenir. Her veri örneklemi için o özelliği taşıyor olma durumuna göre "1" veya "0" değerleri atanır. Şekil 3.1'de de görüldüğü üzere Şehirler "0" ve "1" cinsinden ifade edilmiştir. Bu sayede sayısal olmayan veriler kendi arasında bir ast üst ilişkisi, büyük küçük ilişkisi ifade etmiyorsa uygun şekilde sayısal uzaya taşınmış olur. Bu yöntemdeki en büyük ödün verinin kendi içerisindeki bahsedilen büyük küçük ilişkisi dışındaki bağlamsal ilişkileri göz ardı etmesidir.

ŞEHİR
İstanbul
Ankara
Adana
Ankara



ISTANBUL	ADANA	ANKARA
1	0	0
0	0	1
0	1	0
0	0	1

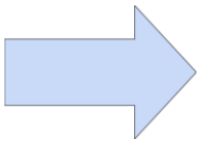
Şekil 3.1- One Hot Encoding Yapısı

3.2. Discrete Encoding

Literatürde Label Encoding olarak geçen Discrete Encoding yöntemi de sayısal olmayan verileri sayısal uzaya taşımak için kullanılan bir yöntemdir. Bu yöntemde her bir sayısal olmayan değere bir tam sayı atanır. Bu sayede sayısal olmayan veri sayısal uzayda ifade edilmiş olur.

Label encoding yönteminde eldeki sayısal olmayan değerler sırası ile bir tam sayı ile ifade edilir. Bu yöntemin Label encoding olarak adlandırılmasındaki temel sebep, ekseriyetle bir veri kümesindeki sayısal olmayan etiketler için uygulanmasıdır. Yöntem uygulanabilirlik açısından kolay ve hızlı olmasına karşın dezavantajları da bulunmaktadır. Yöntem ile eldeki sayısal olmayan veri, tam sayı değeri olarak ifade edilir. Bu da esasında aralarında büyüklük-küçüklük gibi ordinal bağıntılar olmayan veri parçalarını buna göre anlamlandırılmış olur. Örnek olarak, Şekil 3.2'de bir veri kümesinde şehir bilgisinin metin olarak ifade edilmesi ve bunun Discrete Encoding ile dönüştürülmesi ifade edilmiştir. Burada varsayımsal olarak 'İstanbul' ve 'Ankara' arasında ordinal bir bağıntı yoktur. Yapılan bu çalışmada verinin boyutu ve başarım kıstasları değerlendirildiğinde Discrete Encoding yöntemi kullanılmıştır.

ŞEHİR
İstanbul
Ankara
Adana
Ankara



ŞEHİR
0
1
2
1

Şekil 3.2-Discrete Encoding Yapısı

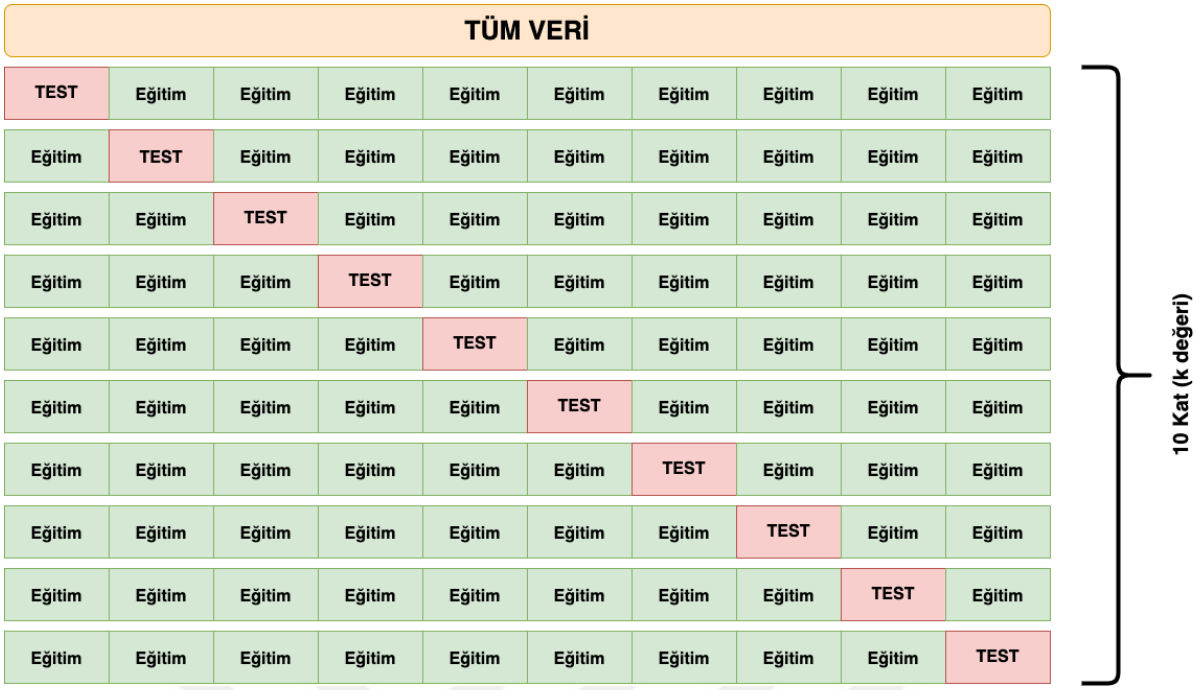
3.3. Normalizasyon İşlemi

Normalizasyon veri işleme safhasını daimî bir parçasıdır. Daimî olması olmasının temel sebebi veri işlemede kolay uygulanabilir ve veri içerisindeki tutarlılığı arttıran bir işlem veya işlemler bütünü olmasıdır. Tanımı gereği normalizasyon, bir veri veya veri kümesi içerisindeki değerlerinin tutarlılığını arttırmaktır. Bu işlem birkaç farklı koldan gerçekleştirilir.

1. **Çoklama Temizliği**: Veri içerisindeki birbirinin kopyası olan örneklerin bulunarak temizlenmesidir.
2. **Aykırı Değer Temizliği**: Veri içerisinde çok yüksek, çok alçak veya anlamsız ve uygunsuz değerler bulunabilir. Örneğin bir satış verisinde anormal bir durum çerçevesinde bir anlık satış değeri çok yüksek gelmiş olabilir veya tam tersi. Ek olarak veri okurken veya yazarken teknik bir aksaklık sonucu veya kazara bir değer bozulabilir. Aykırı Değer tespiti yapılarak bu sorunlar giderilir.
3. **Değer Normalizasyonu**: Literatürde normalizasyonun kendisi olarak geçmektedir. Burada çeşitli yöntemler ile veri içerisinde sayısal değerlerin, örneklem uzayında birbirine yaklaştırılmaları ve modellerin öğrenimim kolaylaştırılması hatta mümkünleştirilmesi hedeflenir.

3.4. Çapraz Doğrulama (Cross Validation) (CV)

Çapraz doğrulama yöntemi geliştirilen modeller için bir doğrulama mekanizmasıdır. Basit fakat etkili bir yöntem ile verinin objektif olarak model üzerinde test edilmesine imkân vermektedir. Birçok çapraz doğrulama yöntemi olmasına karşın temelde hepsi verinin bölünmesine üzerinden uygulanmaktadır. Yapılan çalışmada ilk aşamada Holdout Çapraz Doğrulama uygulanmıştır. Bu yöntemde veri kümesi belirli oranlara göre iki parçaya ayrılmaktadır. Literatürde ekseriyetle 70'e 30 veya 80'e 20 şeklinde uygulanmaktadır. Bu yöntem ile alınan başarı sonrada 10 katlı çapraz doğrulama yöntemi ile onanmıştır. Bu yöntemde Şekil 3.3'te de görüldüğü üzere veri on eşit parçaya bölünür. Arasından ilk parça test ve kalan 9'u eğitim için kullanılır. Tüm paralar için test ve eğitim kümeleri kullanılır. Son adımda çıkan sonuçların ortalaması alınarak ortalama başarım objektif olarak elde edilmiş.



Şekil 3.3- Kullanılan Çapraz Doğrulama Modeli

3.5. Karmaşıklık Matrisi (Confusion Matrix) (CM)

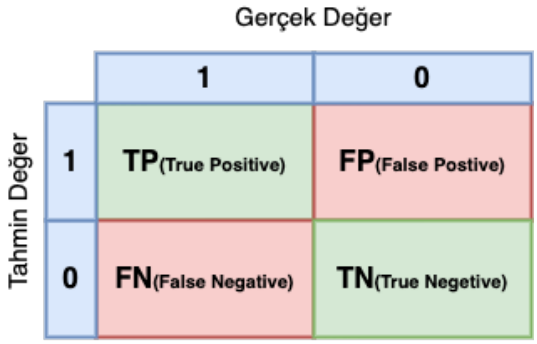
Karmaşıklık matrisi bir modelin başarısını ölçmenin en çok kullanılan yöntemidir. Doğruluk, kesinlik gibi başarım ölçütlerinin temel unsuru olan değerlerin ifade edildiği matristir, Şekil 3.4'te detaylı olarak görülebilir. Bu değerler aşağıda belirtilmiştir. İkili sınıflandırma problemi bahsedilen matrisin anlatımı ve gösterimi açısından kolaylık arz etmektedir. Bu sebeple gösterimde o tercih edilmiştir.

• **Gerçek Pozitif (TP):** Modelin verdiği pozitif (1) tahmin değeri ile, test verisindeki etiket değerinin aynı olduğu.

• **Yanlış Negatif (FN):** Modelin verdiği pozitif (0) tahmin değeri ile, test verisindeki etiket değerinden farklı olduğu.

• **Yanlış Pozitif (FP):** Modelin verdiği pozitif (1) tahmin değeri ile, test verisindeki etiket değerinden farklı olduğu.

• **Gerçek Negatif (TN):** Modelin verdiği pozitif (0) tahmin değeri ile, test verisindeki etiket değerinin aynı olduğu.



Şekil 3.4- Karmaşıklık Matrisi Yapısı

Karmaşıklık matrisi ile mevcut veri kümesinin; doğruluk (accuracy), kesinlik (precision) ve duyarlılık (recall) değerleri hesaplanmıştır. Doğruluk değeri Denklem 3.1'de, kesinlik değeri Denklem 3.2'de, duyarlılık değeri Denklem 3.3'de ve son olarak F1 skoru Denklem 3.4'deki formüller ile hesaplanmıştır.

$$Doğruluk\ (Accuracy) = \frac{TP + TN}{TP + TN + FN + FN} \tag{3.1}$$

$$Kesinlik\ (Precision) = \frac{TP}{TP + FP} \tag{3.2}$$

$$Duyarlılık\ (Recall) = \frac{TP}{TP + FN} \tag{3.3}$$

$$F1 - Skoru = \frac{(2 * kesinlik * recall)}{(kesinlik + duyarlılık)} \tag{3.4}$$

4. ÖNERİLEN MODEL

Yapılan çalışma çerçevesinde, daha önce geliştirilen modellerin başarımlarının üstüne çıkma ve karşılaştırmalı analizler ile sonuçların değerlendirilmesi hedeflenmiştir. Farklı makine öğrenimi yaklaşımı incelenmiş ve sonuçları bahsedilen başarım ölçütleri doğrultusunda değerlendirilmiştir. Bu değerlendirmeler ile en başarı model vurgulanmıştır. Yapılan çalışmada elde veri öz nitelikler bakımından peşi sıra veya zamansal olarak değil, örneklemin içerik özellikleri göz önüne alınarak sonuçlar elde edilmiştir. Bu bölümde önerilen model ve analizler sunulacak ve sonuçları ifade edilecektir. Model geliştirilirken kullanılan akış ifade edilecektir. Ek olarak kullanılan teknoloji ve veri kümesinden de bahsedilecektir.

4.1. Kullanılan Veri Kümesi

Veri kümesi bölümünde detaylı bahsedilen veri kümesi, Hacking and Countermeasures Research Lab araştırma laboratuvarından talep edilmiştir. Veri kümesi farklı parçalardan meydana gelmiştir. DOS saldırısı, bulanık saldırı, vites kutusu yanıltılması ve RPM göstergesinin yanıltılması dahil olmak 4 farklı saldırı mevcuttur. Mesaj enjeksiyon saldırıları yapılırken gerçek bir araçtan OBD-II portu üzerinden CAN trafiği izlenerek veri kümeleri oluşturulmuştur. Her izinsiz giriş 3 ila 5 saniye arasında gerçekleştirilir ve her veri kümesinde toplam 30 ila 40 dakikalık CAN trafiği bulunur.

1. **DOS Saldırısı:** Her 0,3 milisaniyede bir '0000' CAN ID mesajlarının enjekte edilmesi. '0000' en baskın olanıdır.
2. **Bulanık Saldırı:** Her 0,5 milisaniyede bir tamamen rastgele CAN ID ve DATA değerleri içeren mesajlar enjekte etmek.
3. **Yanıltma Saldırısı** (RPM/dişli): Her 1 milisaniyede bir RPM/dişli bilgileriyle ilgili belirli CAN ID mesajlarının enjekte edilmesi.

   Veri seti içerisindeki öz nitelikler aşağıdaki gibidir.

- **Timestamp** : kaydedilen zaman(lar)
- **CAN ID** : HEX'teki CAN mesajının tanımlayıcısı (örn. 043f)
- **DLC** : 0'dan 8'e kadar veri bayt sayısı
- **DATA[0~7]** : veri değeri (bayt)
- **Flag** : T veya R, T enjekte edilen mesajı, R ise normal mesajı temsil eder.

4.2. Modelde Kullanılan Teknoloji

Tez çalışması içi farklı teknolojiler kullanılmıştır. Yapılan araştırmanın verimliliği göz önüne alınarak bu teknoloji tercihleri gerçekleşmiştir. Bu çerçevede geliştirilen tez çalışmasının makine öğrenmesi algoritmalarının ihtiva etmesi, kullanım kolaylığı ve bilgi birikimi gibi sebeplerden ötürü Python programlama dili programlama için tercih edilmiştir. Anaconda paketleri ile çalışma desteklenerek, gerekliliklerin tamamı hızlıca sisteme entegre edilmiştir. Geliştirme temel olarak Jupyter Lab üzerinden gerçekleşmiştir. Jupyter Lab Python kodu parçalarını küçük gruplar halinde çalıştırılmasını sağlayan bir ortamdır. Bu sayede geliştirmeler hızlanır ve sonuçlar hızlıca gözlemlenebilir.

Makine öğrenimi için en zengin paketlerden biri olan Scikit-learn tercih edilmiştir. Verinin okunması ve işlenmesi için ise Pandas paketinden yararlanılmıştır. Pandas ile beraber csv uzantılı dosyaları hızlı ve etkin bir şekilde işlenebilmektedir. Matris operasyonları için numpy kütüphanesinden faydalanılmıştır.

Çok katmanlı algılayıcı modeli için bir derin öğrenme paketi olan Tensorflow tercih edilmiştir. Fonksiyonel ara yüzü ile hızlı ve esnek kullanılma imkân vermektedir. Ayrıca dökümantasyon zenginliği ile de öğrenim ve kullanım için uygun bir ortam olarak ifade edilebilir.

Son olarak grafikler için Mathplotlib kütüphanesinden faydalanılmıştır.

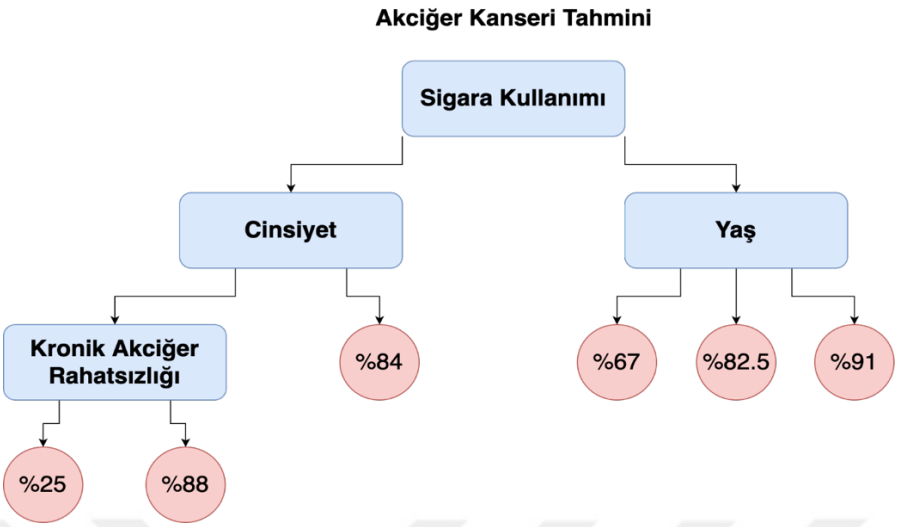
4.3. Kullanılan Algoritmalar

4.3.1. Karar Ağacı Algoritması (DT)

Karar Ağacı yapısı, temel nitelikli bir algoritmadır. Buradaki kasıt Rastgele Orman, Adaptif Artırma, Bagging Ağacı ve daha birçok grup veya bileşke öğrenme metotlarınında temel yapısı olmasıdır. Genel geçer olarak açıklanabilirliği (interpretability) yüksek kabul edilir. Açıklanabilirlikten kasıt algoritmanın öğrenmeyi nasıl gerçekleştirdiği, parametrelerin gözlemlenmesi ve en önemlisi alınan sonucun neden o değer olduğunun mantıksal ve deneysel ispatıdır. Oluşturulmuş bir karar ağacının yapısı Şekil 4.1'deki gibidir.

Öğrenme safhasında verideki öz niteliklerin entropisi hesaplanır. Ardından bu entropi Bilgi Kazanımı veya Gini Index gibi bir hesaplama yöntemi kullanılarak, ağaç dal dal oluşturulur. Ardından oluşturulan ağaçta çıkarım yapılacak veri örneğinin özniteliklerinin durumları "*if else*" yapısı ile hesaplanır. Son noktada ulaşılan yaprak düğüm çıkarımı oluşturur.

Genellikle Sınıflandırma için kullanılan bu yöntem, Regresyon problemlerinde de kullanılabilmektedir. Fakat oradaki en önemli nokta oluşturulan model sürekli değerler üretemez.



Şekil 4.1- Karar Ağacı Yapısı

4.3.2. Naif Bayes Algoritması (NB)

NB sınıflandırma algoritması, Matematikçi Thomas Bayes'in Denklem 4.1'de görüleceği üzere ortaya koyduğu teoremi baz alan bir yaklaşımdır. Bayesci olasılık ilkelerini temel alan bir sıra olasılık hesaplaması ile istenilen veri noktasının sınıfını belirlemeyi amaçlamaktadır.

$$P(A|B) = \frac{\big(P(B|A) * P(A)\big)}{P(B)} \tag{4.1}$$

P (A/B) = Öncül B olayı gerçekleştiğinde A olayının olasılığıdır.

P (A) = A olayı olasılığıdır.

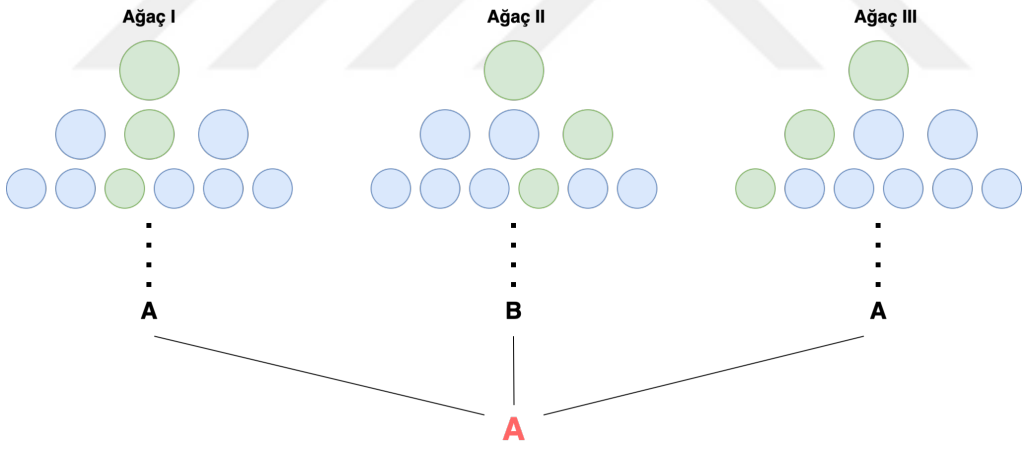
P (B/A) = Öncül A olayı gerçekleştiğinde B olayının olasılığıdır.

P (B) = B olayı olasılığıdır.

Algoritmanın, bir veri noktasının sınıfsal olasılığını hesaplar ve olasılık değeri en yüksek olan sınıf o nokta için tahmin edilen sınıftır. Olasılık temelli olduğu için tüm sınıflar toplamı 1 olmaktadır. Olasılık temelli bir algoritma olması sebebiyle küçük bir veri kümesi ile göreli olarak yeterli sonuçlar verebilmektedir. Tabii olarak verinin artması olasılıkların osilasyonunu azaltacağından dolayı başarım artacaktır. (Doğada var olan örüntülerin ekseri olarak normal dağılıma uygun hareket ettiği varsayılarak). Test verisindeki bir olasılık eğer hiç gözlemlenmemiş ise algoritma olasılık değeri olarak sıfır verir. Bu durum genellikle "Sıfır Frekans" adıyla bilinir. En basit yöntemle tüm veri noktaları için değerler bir artırarak en küçük gözlemin sıfırdan farklı olması saplanarak sorun çözülebilmektedir.

4.3.3. Rastgele Orman Algoritması (RO)

Topluluk temelli algoritmalardan bir tanesidir. Oluşturulan sınıflandırıcılar topluluğunun Karar ağacı türevi olması sebebi ile Orman olarak adlandırılmaktadır. Şekilde 4.2'de de gösterildiği gibi algoritma 'n' tane karar ağacı oluşturur. Oluşturulan bu ağaçlar ile tahmin işlemi yaparak, regresyon veya sınıflandırma için ortalama alma veya çoğunluk oylaması yöntemleri kullanılır. Bu açıklamadan yola çıkarak 'n' değerinin tek sayı olarak tercih edilmesi modelin sınıflandırma yaparken eşit oy probleminin ortaya çıkmasını engeller. Her algoritmada olduğu gibi bu algoritmanın da kendi içerisinde avantajları ve dezavantajları bulunmaktadır. En büyük avantajı bahsedildiği gibi verinin farklı kısımları ile eğitilen birden çok ağaç olduğu için ezberleme (overfitting) problemi azalmaktadır. Esasen topluluk temelli modellerin genel avantajı olarak addedilebilir. Rasgele Orman Algoritmasının karar ağacından farklı olarak işlem gücü dezavantajı bulunmaktadır. Oluşturulan ağaçların sayısı, ağaçların derinlikleri ve budama gibi işlemler bu algoritmalarının işlem temelli gerekliliklerini ver karmaşıklığını arttırmaktır. Her ağaç temelli yaklaşımda olduğu gibi burada da süreklilik arz eden tahminlerde gerçeğe yaklaşmakta diğer algoritmaları göre yapısı daha zayıf kalabilmektedir.

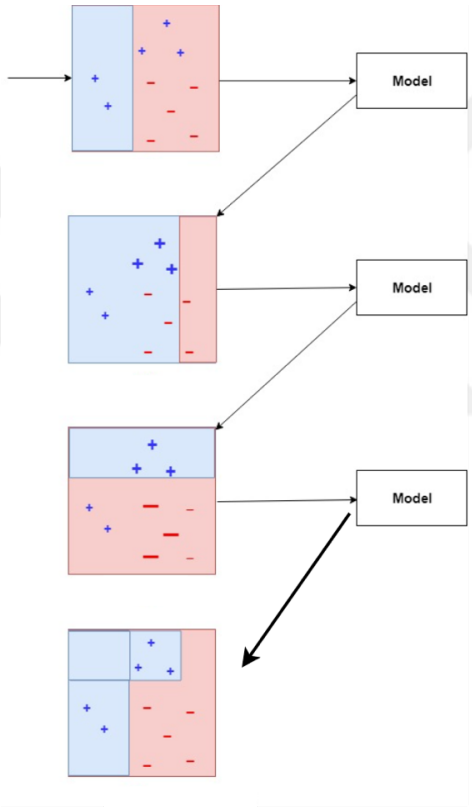


Şekil 4.2- Rastgele Orman Algoritması

4.3.4. Adaptif Arttırma (AA)

AA (Adaboost) algoritması [13] 1996 yılında önerilmiştir. AA dayandığı temel metodoloji, ara modellerin (zayıf sınıflandırıcı) katsayılarını iyileştirmek ve her bir döngüde veri örneğini, o ara modelin olağan dışı olarak nitelediği veri örneklemlerinin doğru tahminlerini sağlayacak şekilde eğitmektir. Başarımı artırmak ve asıl olan sınıflandırıcıyı elde etmek için birçok ara sınıflandırıcıyı birleştirir. Ayrıca AA, diğer eğitim döngülerini iyileştirmek için bir eğitim döngüsünün çıktısını kullanan yinelemeli bir topluluk yöntemidir. AA Sınıflandırıcısı

oluşturabilmek için, sınıflandırıcı eğitim verisi üzerinde eğitilir ve eğitilmiş algoritma çalıştırılarak tahmin yapılır. Daha sonra, hatalı olarak sınıflandırılan verilerin göreceli katsayıları artırılır. İkinci sınıflandırıcı, bu arttırılmış ağırlıklardan faydalanarak eğitilir ardından tekrar tahmin işlemi gerçekleştirilir. İşlemler belirli bir kıstasa kadar bu şekilde devam eder. Algoritmanın sınıflandırıcıyı eğiterek sınıflandırma sınırlarını oluşturması Şekil 4.3'te ifade edilmiştir. Bu kıstas döngü sayısı, başarım vs. olabilir. Yapılan çalışmadaki problem uzayı için 50 farklı karar ağacı yapısı ile AA denenmiştir. Arttırma algoritması olarak SAMME kullanılmıştır [14].

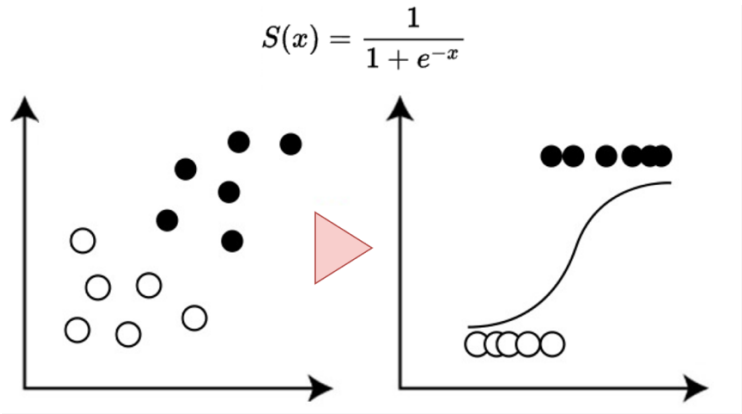


Şekil 4.3-Adaptif Arttırma Yapısı

4.3.5. Lojistik Regresyon (LR)

LR, olasılık problemlerinin üstesinden gelmek için geliştirilmiş istatistiksel bir yaklaşımdır. LR modeli, verilerin desenini öğrenmek için sigmoid işlevi olarak da adlandırılan lojistik işlev kullanır. Sigmoid fonksiyonun yapısı Şekil 4.4 'da görülebilir. Ek olarak, lojistik regresyon modeli, 0 ile 1 arasında bir doğrusal denklemin çıktısını sıkıştırmak için lojistik işlevi kullanılabilir. Regresyon problemlerinin yanında, LR sınıflandırma problemleri için de kullanılabilir. Eğitim aşamasının nasıl yapılacağını tanımlayan bazı parametreler vasıtası ile

modeli eğitmek mümkündür. Geliştirilen modelin ceza seçeneği için L2 normalizasyonu optimize etmek için sınırlı hafızalı BFGS kullanıldı.



Şekil 4.4-Lojistik Regresyon

4.3.6. Örnekleme (Bagging) Algoritması (BA)

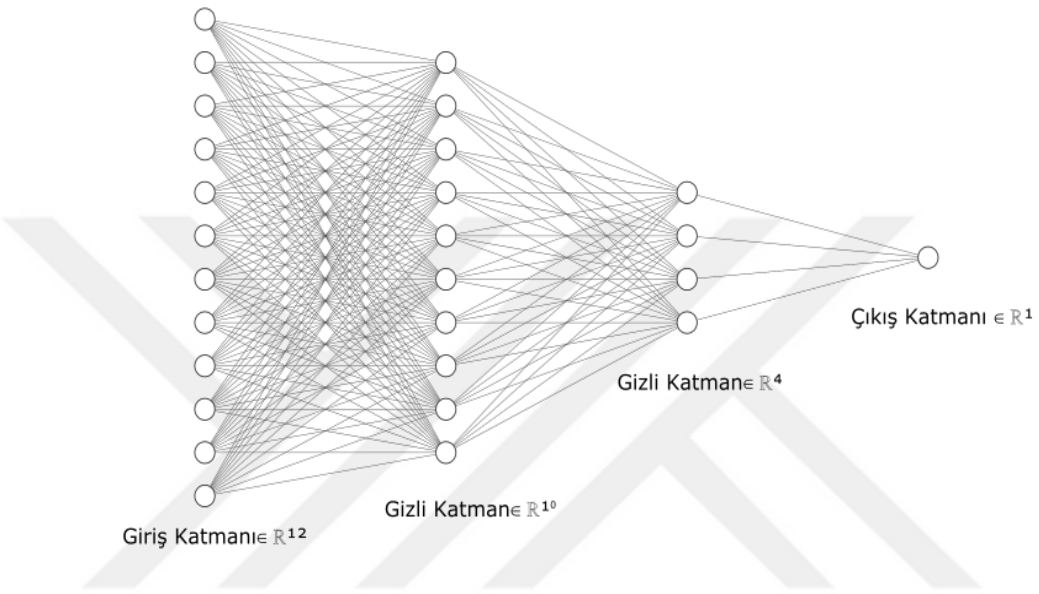
Mevcut bir eğitim setinden yeni eğitim setleri türeterek temel öğreneni yeniden eğitmeyi amaçlayan bir yöntemdir. Örnekleme, yerine koyarak yapılır. Esasen sistemin öğrenimi güçlendiren unsurun temel buradan gelmektedir. Eğitim seti, n örnekten oluşan eğitim setinden m örnekle değiştirilerek rastgele seçim yapılarak üretilir. Seçilen her veri örneği eğitim setine tekrar yerleştirilir. Rastgele olarak örneklemler yeni eğitim setine dahil edilmezken, diğerleri birden fazla defa, tekrar edecek şekilde dahi edilir. Bahsedildiği üzere bu geri koyma işleme modelin overfit eğilimi arttırıyormuş gibi görünmesine karşın verinin öznitelikleri içerisindeki örüntüleri keşfetmeye yarar. Bu sayede birden çok ağaç eğitilir ve kolektif başarı elde edilir.

4.3.7. Çok Katmanlı Algılayıcı (MLP)

Günümüzde derin öğrenme gitgide popülerleşmekte ve gelen bu popülerlik ile birlikte Yapay Sinir Ağı gibi söylevler ön plana çıkmaktadır. Esasen Çok Katmanlı Algılayıcı tabiri basit yahut sığı Yapay Sinir Ağına tekabül etmektedir. Yapay sinir ağı olarak tabir edilen yapı Şekil 4.5'te görüleceği üzere temelde tek yönlü bir hesaplama çizgesidir. Araştırma alanı özelinde bu tip ağlara İleri Beslemeli Yapay Sinir Ağı denilmektedir. Bu basit ağı temel unsur (düğüm) algılayıcıdır.

Algılayıcılar esasen basit yapılı çizge düğümleridir. İki unsurdan meydana gelirler. Ağırlıklı toplam ve aktivasyon fonksiyonu. İlk aşama ağırlıklı toplam olarak adlandırılmaktadır. Bu aşamada diğer algılayıcılardan gelen katsayı ve öznitelikler toplanırlar. Bu toplama işleminden

sonra ikinci aşamada buradan gelen değer bir aktivasyon, eşik değer fonksiyonundan geçerek bir sonraki katmana geçip geçmeyeceği belirlenir. Buradaki en önemli unsur fonksiyonun problem uzayına uygun olması ve her noktasında türevlenebilir olmasıdır. Bu çerçevede çeşit fonksiyonlar geliştirilmiştir. Bunların belli başlı olanları ReLU, Tanh, Sigmoid ve Softmax olarak kullanılmaktadır. Buna ek olarak bu fonksiyonların araştırılması da ayrı bir çalışma alanıdır.



Şekil 4.5- Çok Katmanlı Algılayıcı Yapısı

5. TEST SONUÇLARI VE DEĞERLENDİRMELER

Modeller, sırasıyla birleşik veri kümesi ve ayrı ayrı saldırı kümeleri üzerinde eğitilmiştir. Kaynaktan elde edilen 4 farklı saldırı tipine ait veri setleri birbiri ardına eklenerek tek bir veri setine dönüştürülmüştür. Bu birleşik veri seti 16,5 milyon örnek içermektedir. Modeller daha sonra bu veri seti üzerinde çalıştırılmıştır. Çalıştırılan modellerin sonuçları ayrı ayrı değerlendirilerek sunulmuş ve çıkarımlar gerçekleştirilmiştir.

Sistem performansının hesaplanması için bazı geleneksel metrikler kullanıldı. Bu metriklerden önceki bölümde bahsedilmiştir.

Tüm Veri Kümesi

Elde edilen sonuçlara göre özellikle karar ağacına dayalı algoritmaların daha yüksek başarı elde ettiği gözlemlenmiştir. Ayrıca Tablo 5.1'den de görüleceği üzere özellikle topluluk öğrenme yöntemleri karar ağaçları ile birleştirilmiş, etkili sonuçlar ortaya konulmuştur. Ayrıca, bu modellerin yorumlanabilirliği yüksektir. Bu, modellerin kolayca açıklanabileceği anlamına gelir.

Tablo 5.1-Tüm Veri Kümesi ile Sonuçlar

Model/Score	*Doğruluk*	*Kesinlik*	*Duyarlılık*	*F1- Skoru*
Random Forest	**0,97**	**0,90**	**0,99**	**0,94**
Bagging	**0,97**	**0,90**	**0,99**	**0,94**
Ada Boosting	0,81	0,41	0,53	0,46
Naive Bayes	0,96	0,90	0,94	0,92
Logistic Regression	0,96	0,79	0,76	0,77
Neural Network	0,96	0,78	0,78	0,78

Yanıltma Saldırısı RPM Sonuçları

İncelenen ilk saldırı türü Yanıltma Saldırısı'dır. Bu saldırı RPM Göstergesi üzerinde gerçekleşmiştir. Tablo 5.2'de görüldüğü gibi tespit oranları 1'dir. Buradaki en önemli faktör, mesajların enjekte edilmesinin aynı değerleri içermesidir. Saldırı olarak kabul edilen veri örneklerinin tim öz nitelikleri aynı hexadecimal değerleri içermektedir. Bu nedenle, modeller bunu kolayca öğrenebilir. Bu sebepler modeller, ayrılmış veri kümesi üzerinde tam başarım sergilemiştir.

Tablo 5.2- Yanıltma Saldırısı RPM

Model/Score	*Doğruluk*	*Kesinlik*	*Duyarlılık*	*F1- Skoru*
Random Forest	1,00	1,00	1,00	1,00
Bagging	1,00	1,00	1,00	1,00
Ada Boosting	1,00	1,00	1,00	1,00
Naive Bayes	1,00	1,00	1,00	1,00
Logistic Regression	1,00	1,00	1,00	1,00
Neural Network	1,00	1,00	1,00	1,00

Bulanık Saldırı Sonuçları

İkinci saldırı türü Bulanık saldırıdır. Veri noktaları rastgele değerler içermektedir. Bu saldırı ile birlikte, saldırıya uğrayan sensor ve düğümler hatalı değer üretmeye başlar. Bu rasgelelik durumu, saldırıyı tespit etmeyi zorlaştıran temel unsurlardan biridir. Tablo 5.3'te görülebileceği gibi, modellerin algılama performansı farklılık göstermektedir. Burada dikkat edilmesi gereken en önemli unsur karar ağacı temelli modellerin yüksek başarım göstermesidir. Burada verinin cinsi ve dağılımı büyük önem arz etmektedir. Verinin ayrık değerlerden oluşması, karar ağaçları tarafından verinin öğrenilmesini daha kolay hale getirmektedir.

Tablo 5.3-Bulanık Saldırı

Model/Score	*Doğruluk*	*Kesinlik*	*Duyarlılık*	*F1- Skoru*
Random Forest	**0,99**	**0,99**	**0,99**	**0,99**
Bagging	**0,99**	**0,99**	**0,99**	**0,99**
Ada Boosting	**0,99**	**0,99**	**0,99**	**0,99**
Naive Bayes	0,93	0,93	0,93	0,93
Logistic Regression	0,95	0,95	0,95	0,95
Neural Network	0,94	0,94	0,94	0,94

Yanıltma Saldırısı Gear Saldırısı

Yanıltma Saldırısı Gear Saldırısı bir önceki gibi bulanık saldırı ile aynı şekilde işlemektedir. Buradaki saldırı aracın vites geçiş sensörlerini hedef almaktadır. Bu saldırı tipinde, belirli olan sensörlere aynı veri sıklık ile gönderilerek sensörün çalışması engellenmekte veya üretilen değerler ile bozulmaktadır.

Bahsedildiği üzere saldırı verisi içindeki veri aynı değerlerden ve belirli sensörlere olduğu için model sabit ve yüksek başarım göstermektedir. Önceki saldırı tipinden farklı olarak burada başarımın tam oluşmasındaki temel etken aynı sensör için aynı veri gönderimi doğal veri akışında da sağlanmış olması ve sistemin bunu saldırı olarak değerlendirmesi, dolayısı ise yanlış pozitif değer vermesidir.

Tablo 5.4 Yanıltma Saldırısı

Model/Score	***Doğruluk***	***Kesinlik***	***Duyarlılık***	***F1- Skoru***
Random Forest	1,00	1,00	1,00	1,00
Bagging	1,00	1,00	1,00	1,00
Ada Boosting	1,00	1,00	1,00	1,00
Naive Bayes	1,00	1,00	1,00	1,00
Logistic Regression	1,00	1,00	1,00	1,00
Neural Network	1,00	1,00	1,00	1,00

DOS Saldırısı

DOS saldırı türünde verilerin değeri sabit (‘0000’) olduğundan ve aynı sadece belirli sensör ve aktüatürlere iletildiğinden dolayı başarıların belirlenmesinde sabit değerler gözlenir. Tablo 5.5 görüldüğü üzere DOS saldırı tipi için değerin sabit olduğu ve 1'den farklı olduğu görülmüştür. Bu durumun sebebi, veri kümesinde DOS saldırısını içeren veri noktalarının 0'dan oluşmasıdır. Fakat, 23919 veri noktasının normal olarak etiketlenmiş olsa da veri örnekleminin öz nitelikleri içeriği 0'dır. Bu durum ise sonuçlarda 0,2 sapmaya neden olduğu gözlemlenmiştir.

Tablo 5.5-DOS Saldırısı

Model/Score	***Doğruluk***	***Kesinlik***	***Duyarlılık***	***F1- Skoru***
Random Forest	0,98	0,98	0,98	0,98
Bagging	0,98	0,98	0,98	0,98
Ada Boosting	0,98	0,98	0,98	0,98
Naive Bayes	0,98	0,98	0,98	0,98
Logistic Regression	0,98	0,98	0,98	0,98
Neural Network	0,98	0,98	0,98	0,98

6. SONUÇLAR

Son yıllarda otomobillerin dijitalleşmesi, özellikle güvenlik açısından bazı ek hizmet kalitesiyle kullanıcıların ihtiyaçlarını artırıyor. Kullanılan gömülü sistemler veya bilgi işlem cihazları dijital saldırılara karşı korunmalıdır. Bu nedenle bu yazıda, araçların CAN taşıyıcı yapısındaki saldırıları tespit etmeyi amaçlayan bir Saldırı Tespit Sistemi önerilmiştir. Sistemin uygulanması için 6 farklı öğrenme modeli ile makine öğrenmesi tabanlı bir yaklaşım tercih edilmiştir. Burada bu algoritmalar karşılaştırmalı olarak analiz edilerek sonuçlar değerlendirilmiştir.

Deneysel sonuçlar, ağaç tabanlı ve topluluk öğrenmeyi içeren algoritmaların daha başarılı olduğunu göstermiştir. Bunun temel nedeni ise veri içerisindeki ayrıklık durumudur. Temel veri öznitelikleri devamlı değerler olmayıp ayrık ve on altılık (hexadecimal) değerlerdir. Bunu sebebi ise CAN Taşıyıcı mimarisinin bahsedildiği gibi tasarlanmış olmasıdır. Bundan dolayıdır ki karar ağacı temelli algoritmalar yüksek başarım göstermiştir. Bu algoritmalar ayrıca, bağımsız ağaç yapılarından dolayı kolayca paralelleştirilebilir. Ek olarak modellerin tahminleme sürelileri kısadır. Bu da bu yapının tercih edilebilirliğini artırır. Son olarak karar ağacı oluştuktan, daha etkin olarak C ve C++ gibi dillerde kodlanabilir. Bu sayede araç üzerinde düşük işlem kapasiteli entegrelerde çalıştırılabilir. Son olarak açıklanabilirliği yüksek olduğu için de avantajlıdır.

Bundan sonraki çalışmalarda yapay sinir ağlarının etkilerinin daha detaylı incelenmesi ve veri paketlerinin CAN Taşıyıcı üzerindeki zaman akışına bakılarak tekrarlayan sinir ağı mimarilerinin analizinin yapılabilir ve zaman temelli modeller ile öznitelik temelli modeller kıyaslanabilir.

KAYNAKÇA

[1] The Grand Challenge. (n.d.). Defense Advanced Research Projects Agency. Retrieved January 4, 2022, from https://www.darpa.mil/about-us/timeline/-grand-challenge-for-autonomous-vehicles

[2] Lex Fridman, Daniel E. Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Aleksandr Patsekin, Julia Kindelsberger, Li Ding, Sean Seaman, Alea Mehler, Andrew Sipperley, Anthony Pettinato, Bobbie Seppelt, Linda Angell, Bruce Mehler, Bryan Reimer, "MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation," in IEEE Access, vol. 7, pp. 102021-102038, 2019, doi: 10.1109/ACCESS.2019.2926040.

[3] O. K. Sahingoz and N. Erdogan, “RUBDES: A Rule Based Distributed Event System,” Computer and Information Sciences - ISCIS 2003 Lecture Notes in Computer Science, pp. 284–291, 2003.

[4] O.K. Sahingoz, and A.C. Sonmez, ”Agent-Based Fault Tolerant Distributed Event System”, COMPUTING AND INFORMATICS, 26(5). Retrieved June 29, 2020, from http://www.cai.sk/ojs/index.php/cai/article/view/321/252, 2012

[5] Y. Tekin and O. K. Sahingoz, ”A Publish/Subscribe messaging system for wireless sensor networks,” 2016 Sixth International Conference on Digital Information and Communication Technology and its Applications (DIC-TAP), Konya, 2016, pp. 171-176, doi: 10.1109/DICTAP.2016.7544022.

[6] H. Cam, O. K. Sahingoz, and A. C. Sonmez, “Wireless Sensor Networks Based on Publish/Subscribe Messaging Paradigms,” Advances in Grid and Pervasive Computing Lecture Notes in Computer Science, pp. 233–242, 2011.

[7] O. Zorlu and O. K. Sahingoz, ”Increasing the coverage of homogeneous wireless sensor network by genetic algorithm based deployment,” 2016 Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Konya, 2016, pp. 109-114, doi: 10.1109/DICTAP.2016.7544010.

[8] V2X or Vehicle-to-Everything (2021 Update). (2021, June 20). Thales. Retrieved January 4, 2022, from https://www.thalesgroup.com/en/markets/digital-identity-and-security/iot/industries/automotive/use-cases/v2x

[9] United States. National Highway Traffic Safety Administration. (2006). This is NHTSA : people saving people. Washington, D.C. :U.S. Dept. of Transportation, National Highway Traffic Safety Administration.

[10] E. Seo, H. M. Song and H. K. Kim, ”GIDS: GAN based Intrusion Detection System for In-Vehicle Network,” 2018 16th Annual Conference on Privacy, Security and Trust (PST), Belfast, 2018, pp. 1-6.

[11] I. J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680.

[12] H. M. Song, Jiyoung Woo, Huy Kang Kim, In-vehicle network intrusion detection using deep convolutional neural network, Vehicular Communications, Volume 21, 2020, 100198

[13] Y: Freund, and R.E. Schapire, "Experiments with a New Boosting Algorithm". ICML, 1996.

[14] J. Zhu, S. Rosset, H. Zou, T. Hastie, "Multi-class AdaBoost. Statistics and its interface" 2. 10.4310/SII.2009.v2.n3.a8., 2006.

[15] A. Alshammari, M. Zohdy, D. Debnath and G. Corser, "Classification Approach for Intrusion Detection in Vehicle Systems. Wireless Engineering and Technology" pp: 79-94. 10.4236/wet.2018.94007, 2018.

[16] G. Marchetti, Mirco Stabili, Dario. (2017). Anomaly detection of CAN bus messages through analysis of ID sequences. 1577-1583. 10.1109/IVS.2017.7995934.

[17] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory". Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735, 1997.

[18] P. Baldi. 2011. Autoencoders, unsupervised learning and deep architectures. In Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning workshop - Volume 27 (UTLW'11). JMLR.org, 37–50.

[19] M. Hanselmann, T. Strauss, K. Dormann and H. Ulmer, " CANet: An Unsupervised Intrusion Detection System for High Dimensional CAN Bus Data", IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2982544, 2020.

[20] C. Young, J. Zambreno, H. Olufowobi and G. Bloom "Survey of Automotive Controller Area Network Intrusion Detection Systems". IEEE Design Test. PP. 1-1. 10.1109/MDAT.2019.2899062, 2019.

[21] ocslab.hksecurity.net. 2020. Car-Hacking Dataset - Hacking And Countermeasure Research Lab. [online] Available at: ¡http://ocslab.hksecurity.net/Datasets/CAN-intrusion-dataset¿

[22] S. M. Kouchak and A. Gaffar, "Determinism in future cars: Why autonomous trucks are easier to design," 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2017, pp. 1-6, doi: 10.1109/UIC-ATC.2017.8397598.

[23] Q. Xu, B. Wang, F. Zhang, D. S. Regani, F. Wang and K. J. R. Liu, "Wireless AI in Smart Car: How Smart a Car Can Be?," in IEEE Access, vol. 8, pp. 55091-55112, 2020, doi: 10.1109/ACCESS.2020.2978531.