ÖNERİ SİSTEMLERİ İÇİN GÜRBÜZ TAHMİN MODELLERİNİN GELİŞTİRİLMESİ

Doktora Tezi

Halil İbrahim AYAZ

Eskişehir 2023

ÖNERİ SİSTEMLERİ İÇİN GÜRBÜZ TAHMİN MODELLERİNİN GELİŞTİRİLMESİ

Halil İbrahim AYAZ

Doktora Tezi

Endüstri Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Zehra KAMIŞLI ÖZTÜRK

Eskişehir

Eskişehir Teknik Üniversitesi

Lisansüstü Eğitim Enstitüsü

Ağustos 2023

Bu tez çalışması BAP Komisyonu tarafından kabul edilen 20DRP026 no.lu proje kapsamında desteklenmiştir.

JÜRİ VE ENSTİTÜ ONAYI

Halil İbrahim AYAZ'ın ÖNERİ SİSTEMLERİ İÇİN GÜRBÜZ TAHMİN MODELLERİNİN GELİŞTİRİLMESİ başlıklı çalışması 18/08/2023 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Eskişehir Teknik Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği"nin ilgili maddeleri uyarınca, Endüstri Mühendisliği Anabilim dalında Doktora Tezi olarak kabul edilmiştir.

Jüri Üyeleri	**Unvan Adı Soyadı**	**İmza**
Üye	: Doç. Dr. Zehra KAMIŞLI ÖZTÜRK	
Üye	: Prof. Dr. Onur KAYA	
Üye	: Doç. Dr. Alper BİLGE	
Üye	: Doç. Dr. Emre ÇİMEN	
Üye	: Dr. Öğr. Üyesi Vahit TONGUR	

Prof. Dr. Semra KURAMA

Lisansüstü Eğitim Enstitüsü Müdürü

18/08/2023

DANIŞMAN ONAYI

Danışmanlığını yürüttüğüm Doktora öğrencisi Halil İbrahim AYAZ, ÖNERİ SİSTEMLERİ İÇİN GÜRBÜZ TAHMİN MODELLERİNİN GELİŞTİRİLMESİ başlıklı tez çalışmasını tamamlamıştır. Hazırlamış olduğu tez tarafımca incelenmiş ve öğrencinin tez savunma sınavına alınması bilimsel ve etik açıdan uygun görülmüştür.

Tez Danışmanı

Doç. Dr. Zehra KAMIŞLI ÖZTÜRK

ÖZET

ÖNERİ SİSTEMLERİ İÇİN GÜRBÜZ TAHMİN MODELLERİNİN GELİŞTİRİLMESİ

Halil İbrahim AYAZ

Endüstri Mühendisliği Anabilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Ağustos 2023

Danışman: Doç. Dr. Zehra KAMIŞLI ÖZTÜRK

Öneri sistemleri, aşırı bilgi yükü sorununun ele alınmasında önemli bir rol oynamaktadır. Giderek artan bir literatür öneri sistemlerinin önemini kabul etmektedir. Ancak bu sistemler, öneri sistemlerinin performansını olumsuz yönde etkileyen şilin ataklarına maruz kalmaktadır. Söz konusu saldırıları tespit etmek için birçok çalışma sunulmuş ve başarılı sonuçlar elde edilmiştir. Ancak çalışmaların çoğu çevrimdışı yöntemler sunmuştur. Bu nedenle, bu sistemlerin gürbüzlüğünü artırmak için çevrimiçi ve anlık atak tespit yöntemlerine ihtiyaç duyulmaktadır. Bu tez çalışmasında, öneri sistemlerine anlık müdahaleyi mümkün kılmak için yeni bir çevrimiçi şilin saldırı tespiti önerilmektedir. Şüpheli kullanıcıları tespit etmek için kalite kontrol grafikleri kullanılmıştır. Tespit edilen şüpheli kullanıcılar yoğunluk tabanlı kümeleme algoritması ile kontrol edilmektedir. Kontrol sürecinde kullanıcı benzerliklerini bulmak için korelasyon analizi kullanılmaktadır. Önerilen yöntemin performansını göstermek için duyarlılık, kesinlik ve F1-skorları verilmiştir. Bu çalışmanın ana katkısı, şilin saldırılarının tüm veri kümesi yerine şüpheli kullanıcılar arasında tespit edilmesidir. Ayrıca, saldırı zamanı tespit sürecinde ek bir öznitelik olarak kullanılmıştır. Öneri sistemlerinin gürbüzlüğünü sağlamak için tespit edilen ataklar ile bir sınıflandırma algoritması eğitilerek gelecekteki benzer atakların tespitinin kolaylaştırılması sağlanmıştır. Burada atak ve gerçek profiller arasındaki sayı eşitsizliğini dengelemek amacıyla aşırı örnekleme yöntemi kullanılmıştır. Yine atak profillerin olmadığı durumlarda tek sınıflı destek vektör makinaları ile atak tespiti sağlanmıştır.

Anahtar Sözcükler: Öneri Sistemleri, Gürbüzlük, Şilin Atakları, Çevrimiçi Tespit, Sınıflandırma

ABSTRACT

DEVELOPMENT OF ROBUST PREDICTION MODELS FOR RECOMMENDER SYSTEMS

Halil İbrahim AYAZ

Department of Industrial Engineering

Eskişehir Technical University, Institute of Graduate Programs, August 2023

Supervisor: Assoc. Prof. Dr. Zehra KAMIŞLI ÖZTÜRK

Recommender systems play an important role in addressing the problem of information overload. A growing body of literature recognizes the importance of recommender systems. However, these systems are subject to shilling attacks that adversely affect the performance of recommender systems. Many works have been presented to detect these attacks and have achieved successful results. However, most of the works have presented offline methods. Therefore, there is a need for online and instantaneous attack detection methods to improve the robustness of these systems. In this thesis, we propose a novel online shilling intrusion detection to enable instantaneous intervention in recommender systems. Quality control charts are used to detect suspicious users. The detected suspicious users are checked with a density-based clustering algorithm. Correlation analysis is used to find user similarities in the checking process. Recall, precision and F1-scores are given to show the performance of the proposed method. The main contribution of this work is that shilling attacks are detected among suspicious users instead of the whole dataset. Furthermore, the attack time is used as an additional attribute in the detection process. To ensure the robustness of the recommender systems, a classification algorithm is trained with the detected attacks to facilitate the detection of similar attacks in the future. Here, oversampling is used to compensate for the number disparity between attack and real profiles. Again, in the absence of attack profiles, single class support vector machines were used for attack detection.

Keywords: Recommender Systems, Robustness, Shilling Attacks, Online Detection, Classification

TEŞEKKÜR

Doktora öğrenimim boyunca öğrettikleri, akademiye ve hayata dair tavsiyeleri ile bana her alanda destek olan danışmanım Doç. Dr. Zehra KAMIŞLI ÖZTÜRK'e sonsuz teşekkürlerimi sunuyorum. Her zaman sağladığı destekler için kendisine minnettarım.

Tezim boyunca sağladıkları pozitif eleştiriler sayesinde tezime önemli katkıları olan tez izleme komitesi üyeleri Prof. Dr. Onur KAYA ve Doç. Dr. Alper BİLGE'ye teşekkürlerimi sunarım.

Doktora eğitimimde önemli katkıları olan sağladıkları eğitimler ile bana farklı bakış açıları kazandıran Eskişehir Teknik Üniversitesi Endüstri Mühendisliği bölümü öğretim elemanlarına teşekkür ederim. Doktora öğrenim boyunca desteklerini esirgemeyen Necmettin Erbakan Üniversitesi Endüstri Mühendisliği Bölümü öğretim üyelerine teşekkür ederim.

Hayatım boyunca her an yanımda olan, destekleri ve motivasyonları ile bana sürekli katkı sağlayan aileme teşekkür ederim.

Eskişehir Teknik Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimine tezime olan katkılarından dolayı teşekkür ederim.

Ayrıca doktora eğitimim boyunca sağladıkları burs desteği dolayısıyla TÜBİTAK BİDEB'e teşekkür ederim.

Halil İbrahim AYAZ

18/08/2023

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Teknik Üniversitesi tarafından kullanılan "bilimsel intihal tespit programı"yla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Halil İbrahim AYAZ

İÇİNDEKİLER

Sayfa

TABLOLAR DİZİNİ

Sayfa

ŞEKİLLER DİZİNİ

Sayfa

GÖRSELLER DİZİNİ

Sayfa

SİMGELER VE KISALTMALAR DİZİNİ

I_S	:	Seçilen Ürünler Kümesi (Selected Item)
I_F	:	Dolgu Ürünler Kümesi (Filler Item)
$I_\emptyset$	:	Puanlanmayan Ürünler (Un-rated Item)
I_T	:	Hedef Ürün (Target Item)
QMS	:	Quick Model Selection
ÇKF	:	Çokyüzlü Konik Fonksiyonlar
ÜAHO	:	Üstel Ağırlıklandırılmış Hareketli Ortalama
KT	:	Kümülatif Toplam
SR	:	Shiryaev-Roberts
DP	:	Doğru Pozitif
DN	:	Doğru Negatif
YP	:	Yanlış Pozitif
YN	:	Yanlış Negatif
DVM	:	Destek Vektör Makinaları
RLP	:	Doğrusal Lineer Programlama (Robust Linear Programming)
ODBOT	:	Aykırı Değer Tespiti Tabanlı Aşırı Örnekleme Tekniği (Outlier Detection Based Oversampling Technique)
ROS	:	Rasgele Aşırı Örnekleme (Random Over Sampling)
RA	:	Rastgele Atak
OA	:	Ortalama Atak
MOA	:	Modifiyeli Ortalama Atak
KA	:	Karışık Atak
AO	:	Atak Oranı
DO	:	Dolgu Oranı
DVM	:	Destek Vektör Makineleri
T-DVM	:	Temel Destek Vektör Makineleri
ODBOT-DVM	:	Aykırı Değer Tespitine Dayalı Aşırı Örnekleme Tekniği – Destek Vektör Makineleri
TSDVM	:	Tek Sınıflı Destek Vektör Makineleri
SMOTE	:	Synthetic Minority Oversampling Technique
ROS	:	Rastgele Aşırı Örnekleme (Random OverSampling)

1. GİRİŞ

Günümüzde teknolojinin hızla ilerlemesiyle birlikte, internet kullanımının artması ve dijital platformların yaygınlaşması, kullanıcıların karar verme süreçlerini etkilemiş ve ihtiyaçlarına yönelik doğru ve kişiselleştirilmiş öneriler arayışını beraberinde getirmiştir. Bu durum, öneri sistemlerinin gelişimine önemli bir ivme kazandırmış ve farklı sektörlerde kullanıcı deneyimini zenginleştirmek amacıyla yaygın olarak kullanılan bir araç haline gelmiştir. Öneri sistemleri, doğru zamanda doğru bilgiyi bulmak için İnternet'i kullanmak anlamına gelen aşırı bilgi yükleme sorununun ele alınmasında önemli bir rol oynamaktadır. Gelişen e-ticaret nedeniyle öneri sistemlerine artan bir ilgiden bahsetmek mümkündür. Öneri sistemleri, film ve TV programları (Barragáns-Martínez vd., 2010; Ji vd., 2016) eğitim (Tarus vd., 2018; Zheng, 2018) turizm (Afsahhosseini & Al-Mulla, 2021; Borràs vd., 2014) ve dijital kütüphaneler (Serrano-Guerrero vd., 2011) gibi birçok farklı alan tarafından yaygın olarak kullanılmaktadır.

Öneri sistemleri gerçek dünyada geniş uygulama alanları bulsa da bu sistemler şilin ataklarına karşı savunmasız kalabilmektedir. Bahsedilen ataklar, öneri sistemlerinin mekanizmasına zarar vererek doğruluk oranlarının düşmesine neden olur. Bu ataklar ile hedef ürünün popülerliği, sahte veya kötü niyetli kullanıcılar tarafından artırılır (veya azaltılır). Tablo 1.1'de örnek atak profili verilmiştir. Tablo 1.1 6 filme kullanıcılar tarafından verilen puanları göstermektedir. Burada 1. Kullanıcı için 6. Film hakkındaki puanlaması tahmin edilmek istenirken atak kullanıcılar tarafından altıncı filmin puan ortalaması yükseltilmek istenmektedir. Burada ilk yedi satırdaki profiller gerçek kullanıcı profillerini gösterirken son üç satır ise atak profilleri göstermektedir. Tablo 1.1'den görüleceği üzere atak profiller 6. ürün haricinde puanlamalar yapmıştır. Yapılan bu puanlamalar atağın tespitini zorlaştırmaya yönelik puanlamalardır. Yapılan atak türüne göre farklı puanlamalar mümkündür. Ayrıca bir kullanıcı tarafından tüm ürünlere puanlama yapılması zor olduğu için atak profillerinde de puanlanmayan ürünlere bulunmaktadır. Atak türleri ve hedef ürün haricindeki detaylı bilgiler literatür kısmında verilmiştir. Şilin ataklarını tespit etmek için çok sayıda çalışma sunulmuştur (Bilge vd., 2014; Rezaimehr & Dadkhah, 2021; Si & Li, 2020a). Bahsedilen atakları tespit etmek için sınıflandırma (Batmaz vd., 2020; Zhang & Zhou, 2014) kümeleme (Bhaumik vd., 2011; Zahra vd., 2015; Zhang & Wang, 2020) ve istatistiksel (Bhaumik vd., 2006) tabanlı yaklaşımlar önerilmiştir.

Tablo 1.1. *Örnek atak profilleri*

Kullanıcı	Film1	Film2	Film3	Film4	Film5	Film6
K1	5		4	3	2	?
K2	2		2	3		
K3	5	4	5			2
K4	3	5	4	2	2	1
K5	5	4	4	2	2	2
K6		3		1	2	
K7		4	2	2	3	
A1	4	4	4	3		5
A2	4		3	2		5
A3	4		3		3	5

Bahsi geçen çalışmalar çoğunlukla çevrimdışı yöntemler olup, veri setindeki anormallikler bulunarak tespit süreci sürdürülmektedir. Aynı zamanda literatürde ataklarla ilgili çok sayıda çalışmaya rastlanmak mümkündür. Bilgisayar ve ağ sistemleri uzun süredir benzer ataklara maruz kalmaktadır. Bu sistemler zaman serisi verileri içerir ve ataklar bu veri setinde rastgele görülür. Bu atak modelleri için çeşitli çevrimiçi atak tespit algoritmaları önerilmiş ve bu algoritmalar için yüksek tespit oranları elde edilmiştir. Üretim sistemlerinde olduğu gibi her zaman serisi verisinde de anomaliler görülebilmektedir. Arızalar, değişiklikler ve insan kaynaklı hatalar üretim sürecinde anormalliklere neden olabilir. Söz konusu anormalliklerin anlık olarak kalite kontrol grafikleri ile tespit edildiği çalışmalara rastlanmaktadır. Bu tespitler, hasarı azaltmak için sisteme anında müdahale edilmesini mümkün kılar. Bu tez çalışmasında, yeni bir çevrimiçi şilin atağı tespit mekanizması önerilmiştir. Bu yöntemle sahte kullanıcıların Üstel Ağırlıklandırılmış Hareketli Ortalama (ÜAHO), Kümülatif Toplam (KT) ve Shiryaev-Roberts (SR) gibi kalite kontrol grafikleri ile tespit edilmesi hedeflenmiştir. Bu grafikler, beklenmedik ortalama değişimi durumunda bir sinyal verir. Kalite kontrol grafikleri, sunucularda izinsiz giriş tespiti gibi çevrimiçi atak tespitinde iyi sonuçlar vermektedir. Sonraki adımda, gerçek ve sahte kullanıcıları ayırt etmek için tespit edilen şüpheli kullanıcılara korelasyon tabanlı bir sınıflandırma işlemi uygulanmıştır. Bu sayede herhangi bir atak durumunda anında müdahaleye olanak sağlanmıştır.

Bahsedilen çevrimiçi tespit yöntemlerine alternatif olarak Shiryaev-Roberts prosedürü ile şüpheli kullanıcıların tespiti sağlanmıştır. SR prosedürü literatürde sıklıkla kullanılan alternatiflerine üstün olarak hem varyans hem de ortalama üzerindeki değişimi tek bir grafik üzerinde göstermektedir. Şüpheli kullanıcıların anlık olarak tespitinden

sonra gelecek kullanıcıların sınıfını belirlemek amacı ile çokyüzlü konik fonksiyonlar ile sınıflandırma modeli kurulmuştur. Sınıflar arasında kullanıcı sayısı fazla olduğu için model doğruluk oranını arttırmak için aşırı örnekleme yöntemi uygulanmıştır. Bu sayede modelin doğruluk oranının arttırılması hedeflenmiştir.

Şilin atakları öneri sistemlerine yapılan tek seferlik ataklar olmayabilir. Bu yüzden tespit edilen atakların gelecek süreçte tekrar yapılabileceği göz ardı edilmemelidir. Gelecekte tekrar eden atakların tespiti için tekrar bir tespit yönteminin uygulanması hayli uğraştırıcı ve zaman kaybı bir iştir. Veri kümesine eklenen verilerin sürekli arttığı da göz önüne alınırsa bir süre sonra daha efektif yöntemlere gereksinim duyulacaktır. Bu yüzden gelecek benze atakları etkin bir şekilde tespit edebilmek için bir sınıflandırma algoritmasından yaralanmak kaçınılmazdır. Bu doğrultuda, bu tez çalışmasında tespit edilen ataklar Destek Vektör Makineleri (DVM) modeline verilerek gelecekteki atakların tespiti kolaylaştırılmıştır.

Sınıflandırma algoritmalarının etki çalışabilmesi için sınıflarda yeterli farklılıkta elemanın olması gerekmektedir. Aksi takdirde algoritmanın performansı düşebilir. Şilin atakları da veri kümesinin ufak bir kısmını oluşturduğu için sınıflar arasında büyük bir sayı eşitsizliği bulunmaktadır. Bu durumla başa çıkabilmek için aşırı örnekleme yöntemlerinden yararlanılmaktadır. Burada kullanılacak yöntem hem atak sınıfının sayısını artırırken hem de çeşitliliğini artırmak durumundadır. Bu iki amacı birlikte sağlayabilmek için son zamanlarda önerilmiş Aykırı Değer Tespitine Dayalı Aşırı Örnekleme Tekniği (Outlier Detection-Based Oversampling Technique, ODBOT) kullanılmıştır. Bu yöntem ile ataklardaki çeşitlilik olması durumunda da yüksek performans ölçütleri değerleri elde edilmiştir.

Literatürdeki şilin atak tespit yöntemlerinin çoğunu denetimsiz öğrenme yöntemleri oluşturmaktadır. Bu yöntemlerden olan K-Ortalamalar yöntemi incelendiğin küme merkezlerinin rasgele seçilmesi bazı durumlarda sorunlar oluşturabilmektedir. Bu durumla başa çıkabilmek için matematiksel tabanlı bir gözetimsiz öğrenme yöntemi önerilmiştir. Atakların Tek Sınıflı Destek Vektör Makineleri ile tespit edilmesi amaçlanırken rassallıkların ortadan kaldırılması amaçlanmıştır.

1.1. Öneri Sistemlerine Genel Bakış

Öneri sistemlerinde kullanıcılara daha uygun öneriler sunabilmek için farklı algoritmalar önerilmektedir. Bazı durumlarda ürünlerin bilgilerinin olduğu veri kümeleri

kullanılırken, kullanıcıların ürünlere verdiği skorlardan oluşan veri kümeleri de kullanılabilmektedir. Bahse konu iki veri kümesi farklı bilgiler içerdiği için önerilen algoritmalarda farklılık gösterebilmektedir. Bununla birlikte teknolojinin gelişmesi ile yöntemler melezlenerek, melezlenen yöntemlerin güçlü yanları bir araya getirilerek sunulan önerilerin kalitesinin artırılması hedeflenmektedir. Devam eden kısımda bu yöntemler ile alakalı genel bilgiler verilmiştir.

1.1.1. Ortak Filtreleme (OF) öneri sistemleri

Ortak filtreleme öneri sistemlerinde amaç, sistemde kayıtlı kullanıcıların geçmişteki davranışlarını göz önüne alarak yeni bir kullanıcı için öneriler sunmaktır. Öneri sistemlerinin ortaya atılmasından itibaren kullanılmakta olan bu yöntemler performansları, kısıtları ve avantajları yönlerinden derinlemesine incelenmiştir. İşbirliğine dayalı yöntemlerde kullanıcı–ürün matrisi ele alınır ve bu matris değerlendirilerek öneri listesi oluşturulur.

Öneri sistemleri için kullanılan ilk yaklaşımlardan olan kullanıcı tabanlı en yakın komşu önermesi, aktif kullanıcı ile veri tabanında ki benzer özellikleri en fazla olan kullanıcıyı araştırma üzerine kurulmuş bir yaklaşımdır. Yaklaşım temel olarak iki varsayıma dayanmaktadır:

(a) Kullanıcılar geçmişte aynı zevklere sahipse gelecekte de aynı zevklere sahip olacaktır.

(b) Kullanıcı tercihleri zaman içerisinde tutarlı ve istikrarlıdır. Yöntem iki kullanıcı arasında benzerliğe bakarken Pearson korelasyon katsayını kullanmakta, katsayı [-1,1] arasında değişmektedir.

Komşu seçimlerinde en yüksek pozitif korelasyonlu komşu seçilmektedir. Tüm kullanıcıların komşuluğa dâhil edilmesi zaman açısından ve önerinin doğruluğu açısından problemlere nede olmaktadır. Komşu sayısını azaltmak için en çok kullanılan yöntemlerden biri benzerlik alt limitinin kullanılması ya da sabit bir komşu sayısının kullanılmasıdır.

1.1.2. İçerik tabanlı filtreleme sistemleri

Kullanıcı tabanlı işbirlikçi yaklaşımlar birçok alana başarılı bir şekilde uygulanmasına rağmen, büyük boyutlu problemlerde çok fazla sayıda komşu araştırması ve birçok ürünün araştırılması durumlarının üstesinden gelememektedir. Bundan dolayı,

büyük veri matrisleri ile yapılacak işlemlerde ürün tabanlı öneriler tercih edilmektedir. Bu öneri sistemleri çevrim dışı ön işleme operasyonlarına daha uygundur. Böylelikle büyük boyutlu matrislerde gerçek zamanlı önerilere olanak vermektedir. Ürün tabanlı algoritmalar için genel işleyiş şu şekildedir. İlk olarak tüm kullanıcıların profilleri çıkarılır. Örneğin bir bilgisayar alacak müşteri için bilgisayardan beklentileri ve talep ettiği özellikler sisteme kaydedilir. Ardından ürün listesindeki bilgisayarlardan müşteri beklentisine en çok uyum sağlayan bilgisayarlar müşteriye önerilir.

Kosinüs benzerliği ürün tabanlı öneri yaklaşımları için kabul edilmiş bir metriktir. Metrik, iki *n* boyutlu vektör arasındaki benzerliği aradaki açıyı temel alarak hesaplar. Sonucunda ürün ile kullanıcı arasındaki benzerlik ortaya koyulur.

OF yöntemleri hafızaya dayalı ya da modele dayalı olarak adlandırılmaktadır. Geleneksel kullanıcı tabanlı teknikte puan tablosu veri tabanında tutulur ve önerme işlemi aşamasında buradan çıkarım yapılır. Bu yüzden teknik hafızaya dayalıdır. Modele dayalı yöntemlerde ham veri bir kez çevrimdışı olarak işlenir ve oradan çıkarılan model ile önermeler yapılır.

1.1.3. Melez yöntemler

Melez yöntemler, iki ya da daha fazla yöntemin güçlü yönlerinin bir araya getirilerek büyük sistem optimizasyonu problemlerini çözmek için önerilmiş modeller ortaya koyma isteğidir. Öneri sistemlerindeki amacı ise doğruluk oranı yüksek ve kısa sürelerde önerileri sunabilen modeller elde etmektir.

Melez yöntemler diğer yöntemlerin en iyi özelliklerini problemin belirli kısımlarında kullanarak karşılaşılan bazı güçlükleri ortadan kaldırmayı amaçlar. Öneri sistemlerinde en çok kullanılan melez yöntemler genellikle içerik tabanlı yöntem ile diğer öneri yöntemlerinin bir araya gelmesiyle oluşur ve verilerdeki seyreklik, ölçeklenebilirlik ve başlangıçta karşılaşılan soğuk başlatma problemleri ile başa çıkmak hedeflenir.

1.1.4. Diğer model ve ön işleme tabanlı yaklaşımlar

İş birlikçi yöntemler hafızaya dayalı ya da modele dayalı olarak adlandırılmaktadır. Geleneksel kullanıcı tabanlı teknikte puan tablosu veri tabanında tutulur ve önerme işlemi aşamasında buradan çıkarım yapılır. Bu yüzden teknik hafızaya dayalıdır. Modele dayalı yöntemlerde ham veri bir kez çevrimdışı olarak işlenir ve oradan çıkarılan model ile önermeler yapılır.

Matris faktorizasyon yöntemleri, puan desenlerini kullanarak gizli faktörleri türetir, kullanıcı ve ürünleri de bu faktörlerin vektörleri ile karakterize eder.

1.1.5. Öneri sistemlerinde karşılaşılan zorluklar

Yukarıda bahsedilen yöntemlerin birbirlerine göre üstünlükleri ya da eksiklikleri vardır. Fakat bu yöntemlerinin hepsinin ortak olarak karşılaştıkları güçlükler söz konusudur. Bunlardan ilki kullanıcı tercihlerinin değişmesidir. Kullanıcıların tercihleri zaman içerisinde çevresel ya da farklı etmenler doğrultusunda değişiklik gösterebilmektedir. Bu yüzden kullanıcıya sunulacak önerilerde zaman içerisinde değişiklik göstermek durumunda kalabilir.

Diğer bir önemli güçlük ise başlangıçtaki matriste eksik olan bilgilerdir. Öneri sistemlerinde genel olarak büyük veri setleri ile çalışılmaktadır. Film öneri sistemleri üzerinden değerlendirecek olursak, bir kullanıcının tüm filmler üzerinde değerlendirme yapması mümkün görünmemektedir. Literatürde sıklıkla kullanılan veri setleri %90 üzerinde eksiklikler içermektedir. Eksik veriler modelin doğruluğunu etkilemektedir.

Yöntemlerin karşılaştığı bir diğer sorun ise mahremiyet ile ilgili problemlerdir. Geliştirilen yöntemlerde daha yüksek doğruluk oranlarının elde edilebilmesi için kurulan modele olabildiğince fazla bilgi vermek gerekir. Fakat bahsedilen durum genelde mümkün olmamaktadır. Yasal veya farklı düzenlemeler ile kişilerin tüm bilgilerinin işlenmesi mümkün olmayabilir. Bunlardan dolayı daha etkin modelin kurulması kısıtlanmaktadır.

Kullanıcı sayısının ve ürün sayısının artması ile ölçeklenebilirlik problemi ortaya çıkmaktadır. Bir başka deyişle, veri miktarı arttıkça aynı zamanda aynı hizmeti vermek mümkün olmayabilir. Bu yüzden kurulacak modellerin daha büyük veriler ile çalışabilir ya da kolay entegre edilebilir olması gerekmektedir.

Yukarıda bahsedilen güçlüklerin yanında bazı özel problemler için özel durumlar söz konusu olabilmektedir. Örneğin eş anlamlı kelimeler bilgisayarlar tarafından ayırt edilmesi mümkün olmamaktadır. Kullanıcıya aynı ürünün farklı kelimeler ile yazılmış versiyonları (kot ve denim gibi) tekrar tekrar öneri olarak gelebilmektedir. Aksi durumda ise sadece bir kelimeyi filtreleyip öneriler getirebilmektedir.

1.2. Tezin Amacı ve Katkılar

Bu tez çalışmasında öneri sistemlerinin gürbüzlüğünü sağlamak için şilin atakları incelenmiştir. Literatür incelendiğinde anlık tespit yöntemlerinin nadirliği gözlemlenmiştir. Bu alanı geliştirmek için çalışmada kalite kontrol grafikleri öneri sistemlerinde kullanılan veri kümelerine entegre edilmiştir. Kalite kontrol grafiklerinde tespit edilen şüpheli kullanıcılar DBSCAN yöntemi ile kümelenerek burada atak kullanıcılar tespit edilmiştir. Böylece zaman serisi halindeki verilerde atakların anlık olarak tespiti sağlanmıştır. Yöntemin ilerleyen zamanlarda aynı ataklara maruz kalma durumunda atakların tespitini hızlandırmak amacıyla çokyüzlü konik fonksiyonlar (ÇKF) ve DVM ile sınıflandırma yapılmıştır. Burada sınıflandırma başarısını artırmak için aşırı örnekleme yöntemi kullanılmıştır. Yani, sayısı az olan atak kullanıcılardan ODBOT yöntemi ile sentetik veriler üretilerek atak ve normal kullanıcılar arasındaki fark azaltılarak başarının artırılması hedeflenmiştir. Son olarak literatürde var olan gözetimsiz atak tespit yöntemlerine alternatif TSDVM ile atak tespiti sağlanmıştır.

1.3. Tez Organizasyonu

Birinci bölümde öneri sistemleri hakkında genel bilgiler verilmiş ve öneri sistemleri yöntemlerine değinilmiştir. Ayrıca öneri sistemlerinin maruz kaldığı şilin ataklarından bahsedilmiştir. İkinci bölümde, kalite kontrol grafikleri, öneri sistemleri, şilin atakları ve tespit yöntemleri, DBSCAN algoritması, çokyüzlü konik fonksiyonlar ve aşırı örnekleme yöntemleri üzerine bir literatür araştırması sunulmuştur. Üçüncü bölümde bu tez çalışmasında kullanılan yöntemler açıklanmış ve matematiksel formülasyonları verilmiştir. Dördüncü bölümde, veri kümeleri üzerine uygulamalar yapılarak yöntemlerin uygulanabilirliği gösterilmiştir. Yöntemlerin sonuçları farklı senaryolar halinde test edilmiş sonuçlar tablolar halinde sunulmuştur. Beşinci bölümde gelecekteki benzer atakların tespiti için aşırı örnekleme tabanlı DVM ile atak tespitine yönelik çalışmanın sonuçlar verilmiştir. Altıncı bölümde TSDVM ile atak tespiti çalışmasının amacı, literatürü ve sonuçları sunulmuştur. Yedinci ve son kısımda, tezin sonuçları özetlenerek verilmiş gelecek çalışmalar için öneriler sunulmuştur.

2. LİTERATÜR ÇALIŞMALARI

Bu bölümde, atakların doğasını anlamak için şilin atak türleri hakkında kısa bilgi verilmiştir. Daha sonra mevcut şilin atak tespit yöntemleri incelenmiştir. Ardından farklı sistemler için benzer atak tespit algoritmaları ve yapılan çevrimiçi atak tespit çalışmaları devamındaki alt bölümlerde özetlenmiştir. Çalışmada kullanılan yöntemlerin literatür çalışmaları özet olarak sunulmuştur. Çalışmanın geçerliliğini göstermek amacı ile literatürde kullanılan performans ölçütleri değerlendirilerek hangi ölçütün nerede ve hangi amaçla kullanıldığı örnekler ile açıklanmıştır. Yine bu çalışmada kullanılan ÇKF ve DVM makineleri ile alakalı çalışmalar bu bölümde sunulmuştur.

2.1. Atak Türleri ve Temel Kavramlar

Öneri sistemlerinde atakların tespitini güçleştirmek ve elde olan bilgiye göre farklı ataklar enjekte edilmektedir. Bir başka ismiyle profil enjeksiyonu olarak adlandırılan şilin atakları öneri sisteminin gürbüzlüğüne zarar vermektedir. Öneri sistemlerinde uygulanan atakların büyüklüğü farklılık göstermektedir. Atak büyüklüğü veya boyutu veri kümesine eklenen atak profil miktarı olarak tanımlanabilir. Yine atakların tespitini güçleştirmek için hedef ürün ile birlikte başka ürünlere yapılan puanlamalardan bahsetmek mümkündür. Hedef ürünler ile birlikte yapılan puanlamalar dolgu ürünlerine yapılmaktadır. Bu ürünlerin sayısı da dolgu boyutu ya da büyüklüğü olarak adlandırılmaktadır.

Veri kümesine eklenen atak profiller farklı amaçlar doğrultusunda puanlanmaktadır. Atağın amacı hedef ürünün ortalamasını düşürmeye yönelik ise, bu tür ataklar çekme türü atak olarak adlandırılmaktadır. Aksi durumda, hedef ürünün puan ortalaması yükseltilmek isteniyorsa bu tür ataklarda itme türü atak olarak adlandırılmaktadır.

Bhaumik vd. tarafından tanımlanan bir atak profili Tablo 2.1'de verilmiştir (Bhaumik vd., 2006). Genel olarak bir atak profili 4 kısımdan oluşmaktadır. I_S atağın bir puanlama fonksiyonu ile atağın karakteristiğini belirlediği puan verilen ürünlerdir. I_F ise atağın tespitini güçleştirmek için yapılan hedef ürün haricinde yapılan puanlamalardır. Burada yapılan puanlamalar ortalama dikkate alınarak yapılır ve gerçek kullanıcı gibi davranılır. Öneri sistemlerinde bir kullanıcının tüm ürünlere puan vermesi olası bir durum değildir. Bu yüzden, I_O da atak profillerinde de puanlama yapılmayan ürün kümesidir. I_T

ise atak yapılmak istenen üründür. Burada amaç, hedef ürünün puanının artırılması ya da azaltılmasıdır. Birçok atak türünden bazıları devam eden kısımda özetlenerek verilmiştir.

Tablo 2.1. *Genel atak örneği*

I_S	I_F	I_O	I_T
Atağın karakteristiğini belirleyen puanlamalar	Atağın tespitini zorlaştıran puanlamalar	Puanlama yapılmamış ürünler	Hedef ürün

Temel ataklar I_F ve I_T'den oluşur. Genel olarak, I_F'de atanan puanlar rastgele oluşturulur. Rastgele ve ortalama ataklar, iyi bilinen iki temel atak türüdür. Temel atak profilleri gösterimi Tablo 2.2'de verilmiştir. Rastgele ataklar en basit ataklardır ve hedef ürün haricinde yapılan puanlamalar genel ortalamaya göre puanlanır (Lam & Riedl, 2004). Bu atak türünde, seçilen öğe kümesi boştur, yani $I_S = \emptyset$. Hedeflenen öğeler kümesi saldırı türüne bağlı olarak maksimum ya da minimum puanlar ile puanlanır.

Öte yandan, ortalama ataklar hedef ürün haricinde yapılan ataklarda her ürünün kendi ortalama ve standart sapmasını kullanır(Burke vd., 2005). Bu yüzden gerçek kullanıcılara kıyasla ayırt edilmesi zordur. Hedeflenen öğelerin derecelendirme modeli rastgele saldırıda olduğu gibi aynıdır.

Sürü atakları, rastgele atakların bir uzantısı olarak düşünülebilir. Bu modelde saldırgan, Zipf'in popülerlik yasası dağılımından yararlanır ve en popüler öğeleri içeren taraflı profiller oluşturur. Popüler öğeler, çok sayıda kullanıcı tarafından derecelendirilen öğelerdir. Bu nedenle, saldırganların gerçek kullanıcılara benzeme olasılığı yüksektir. Seçilen ürünler kümesi sık puanlanan öğelerden oluşan bir kümedir, bu nedenle bu öğelere hedef kümedeki öğelerle birlikte maksimum puanlar atanır. Sürü atak profili gösterimi Tablo 2.2'de verilmiştir. Literatürde tanımlanan iki tip sürü atağı vardır: ortalama (Wu vd., 2012) ve rastgele (Yang vd., 2016) sürü atakları. Her iki atak türünde de popüler ürünler seçilir ve hedeflenen ürünle birlikte onlara maksimum puanlar verilir. Rastgele sürü ataklarında hedef ürün harici yapılan puanlamalar rasgele atak türünde olduğu gibidir.

Benzer şekilde, ortalama sürü atak türündeki hedef ürün haricinde yapılan puanlamalar, ortalama atak türüyle aynı şekilde puanlanır. Bölüm atak modeli, ürün tabanlı öneri sistemlerinde etkilidir (Mobasher vd., 2005).

Bölüm atak modeli, hedefte belirli bir ürünü satın almak isteyen belirli bir kullanıcı grubuna yöneliktir. Örneğin, daha önce macera romanlarına yüksek puanlar vermiş kullanıcılara macera romanları önermeyi amaçlamaktadır. Başka bir deyişle, belirli bir ürünü tanıtmak isteyen bir saldırganın, ürünün tüm kullanıcılara ne sıklıkla tavsiye edildiğiyle değil, muhtemel alıcılara ne sıklıkla tavsiye edildiğiyle ilgilenmesi muhtemeldir. Tablo 2.2'de bölüm atak modeli gösterilmiştir. (Mobasher vd., 2005) tarafından tanıtılmış bu atak türü özellikle ürün tabanlı algoritmalarda daha etkili olmaktadır. Saldırı profillerinde saldırganlar, hedef öğelerle birlikte segmentteki kullanıcıların muhtemelen beğendiği öğeler için maksimum derecelendirme değeri ve dolgu öğeleri için minimum derecelendirme değeri ekler, böylece öğe benzerliklerinin varyasyonlarını en üst düzeye çıkarır (Burke, Mobasher, & Bhaumik, 2005).

Yoklayıcı atak modeli, öneri sistemi (Mobasher vd., 2007) tarafından oluşturulan profilleri kullanır. Gerçek puan profilleri sisteme çekirdek profil olarak verilmektedir. Hedef ürünün puan değeri, elde edilen çekirdek profiller kullanılarak değiştirilir. Komşu tabanlı yöntemlere dayalı güçlü bir atak modelidir (O'Mahony vd., 2005). Bir anlamda, yoklayıcı saldırısı, saldırganların sistemin derecelendirme dağılımını aşamalı olarak öğrenmeleri için bir yol sağlar. Bu saldırının basit olmasının yanı sıra bir diğer avantajı da daha az alan bilgisi gerektirmesidir. Burada, bu çalışmalar, ürünün popülerliğini artırmak için atak modelleri incelenmiştir. Öte yandan, literatürde öğenin popülerliğini azaltmak için atak modelleri de bulunmaktadır.

Sevme/nefret atak modeli, hem ürün tabanlı hem de kullanıcı tabanlı filtreleme yöntemlerinde etkili olan atak sırasında bilgi gerektirmeyen atak modelidir (Williams & Mobasher, 2006). Bu atak modelinde doldurulacak ürünler rastgele seçilir ve minimum puanla puanlanırken, hedef atak ürünü maksimum puanla puanlanır. Tablo 2.2'de atak profili özetlenmiştir.

Ters sürü saldırısı belirli ürünleri bombalamak için yukarıda tartışılan sürü saldırısının bir varyasyonudur. Bu saldırıda profiller, hedef öğeler kümesi ile birlikte en az popüler öğelere minimum derecelendirme değeri verilmesine dayalı olarak oluşturulur. Dolgu öğeler için derecelendirmeler rastgele saldırıda olduğu gibi rastgele belirlenir. Benzer şekilde, ters sürü saldırısının uygulanması nispeten kolaydır. Ters sürü atak profili Tablo 2.2'de özetlenmiştir. Ters sürü (ortalama) saldırı ve ters sürü (rastgele) saldırı

olarak adlandırılabilecek iki farklı ters sürü saldırı modeli vardır. Yapılan çalışmalarda görülmüştür ki ürün tabanlı algoritmalarda çok etkilidir (Cheng & Hurley, 2010a).

Tablo 2.2. *Atak profilleri*

Atak Modeli	Seçilen Ürünler		Dolgu Ürünleri		Hedef Ürün
	Ürün	Puan	Ürün	Puan	
Rastgele	X	X	Rastgele	Genel ortalama	r_{max}/r_{min}
Ortalama	X	X	Rastgele	Ürün Ortalaması	r_{max}/r_{min}
Sürü	Popüler ürünler	r_{max}	Rastgele	Genel ortalama	r_{max}/r_{min}
Bölüm	Bölümlenmiş	r_{max}/r_{min}	Rastgele	Genel Ortalama	r_{max}/r_{min}
Sevme/Nefret	X	X	Rastgele	r_{max}	r_{min}
Ters Sürü	Popüler olmayan ürünler	r_{min}	Rastgele	Genel ortalama	r_{min}

Saldırganlar, mükemmel bilgi atak modelinde veri kümesi dağılımlarını yeniden üretebilir (Williams & Mobasher, 2006). Saldırganlar, mükemmel bilgi atak modelinde veri kümesinde bulunan önyargılı profilleri kullanır, lehine veya aleyhine önyargı sergiledikleri bazı belirli öğeler dışında, sistemde halihazırda bulunan profillerle tam olarak eşleşir. Ancak, bu atak modelinin uygulanması zordur çünkü tüm veri kümesinin bilgisini gerektirir.

Örnekleme saldırısı, profillerin gerçek profil veri tabanından örneklenen tüm kullanıcı profillerinden oluşturulduğu ve itilen öğe için olumlu bir derecelendirme ile artırıldığı bu tür bir saldırıdır. Ortak Filtreleme algoritmalarında istikrarsızlığını gösterse de orijinal kullanıcı profillerine kolayca erişilemediği için bu tür bir saldırıyı gerçekleştirmek nispeten pratik değildir (Burke vd., 2005).

2.2. Şilin Ataklarına Karşı Geliştirilen Yöntemler

Şilin atak tespiti, öneri sistemlerinde hayati bir rol oynamaktadır. Bahsedilen atakları tespit etme yöntemleri denetimli, denetimsiz ve yarı denetimli yöntemler olmak üzere üç kısımda incelenebilir. Denetimli yöntemler, şilin ataklarını tespit etmek için etiketli verileri kullanmaktadır. Bir model, atak ve özgün profiller kullanılarak eğitilir ve ardından model doğrulaması sağlanırsa, test kümesindeki profillerin atak olup olmadığı için bu model kullanılır. Denetimsiz yöntemler etiket bilgisi gerektirmemektedir. Ayrıca kümeleme tabanlı modeller olarak da adlandırılırlar. Tüm kullanıcılar farklı kümeleme yöntemleriyle kümelenir. Bu yöntem, atak profilleri arasındaki benzerliğin yüksek olacağı varsayımıyla çalışır. Son kümelerde, yüksek benzerlik oranlarına sahip kümeler

atak profili olarak kabul edilebilir. Yarı denetimli yöntemler, az sayıda etiketlenmiş veri ve birçok etiketlenmemiş veri kullanır. Modeli eğitmek için az sayıda etiketli veri kullanılır.

Denetimli yöntemlerin algılama yeteneği, eğitim verilerinin miktarı ve eğitim verileri üzerindeki atak çeşitliliği ile doğru orantılıdır. Bu nedenle, sınıflandırma tabanlı modeller için farklı modellerin özellikleri kapsamlı bir şekilde çıkarılmalıdır. Bununla birlikte, geleneksel modeller, özellik çıkarma için yetersiz olabilir. (Yang vd., 2016)'da farklı atak modellerinin istatistiksel özellikleri kullanılarak öznitelikler çıkarılmıştır. Ardından, önerilen algoritma, çıkarılan özniteliklere dayalı olarak atak profillerini tespit etmek için kullanılmıştır. En bilindik yöntemlerden biri olan Destek Vektör Makineleri (DVM) yüksek doğruluk oranları verir (Zhang & Zhou, 2014). Önerilen çalışmada önce atak profilleri belirlenmekte ve daha sonra DVM ile sınıflandırılmaktadır.

Denetimsiz yöntemler, atakları tespit etmek için kullanıcılar arasındaki benzerlik veya mesafeyi kullanır. Denetimsiz yöntemler ilk olarak O'Mahony ve arkadaşları tarafından öneri sistemlerinde kullanılmıştır (O'Mahony vd., 2003). Kötü niyetli kullanıcıları tespit etmek ve öneri sistemlerinin sağlamlığını artırmak için kümeleme yöntemleri uygulanır. Öneri sistemlerinde atak tespiti için yaygın bir kümeleme yöntemi olan K-ortalamalar uygulanır (Bhaumik vd., 2011). (Cai & Zhang, 2021)'de, davranış özelliği çıkarma ve davranış matrisi benzerliği, atak ve gerçek kullanıcılar arasında ayrım yapar. Önerilen yöntem, atak ve hedef ürün haricinde yapılan puanlamaları dikkate almadan atak profillerini tespit etmektedir. Diğer açıdan, öneri kalitesini artırmak için bazı yöntemler kümeleme yöntemleriyle hibrittir (ZCheng & Hurley, 2010b; Mehta & Nejdl, 2009).

Yoğunluk tabanlı DBSCAN algoritması rastgele şekiller içeren verilerin kümelenmesi ve bu veriler içindeki gürültülerin tespiti için önerilmiş bir yöntemdir (Sander vd., 1998). Özellikle dışbükey olmayan, küresel ve bir yoğunluk doğrultusunda uzamış olan verilerde iyi sonuçlar veren algoritmanın öneri sistemleri üzerindeki uygulamalarına literatürde az rastlanmaktadır. (Nawara & Kashef, 2021)'de sunulan çalışmada, DBSCAN algoritmasının farklı kümeleme algoritmaları ile karşılaştırmalı sonuçları verilmiştir. (Kużelewska & Wichowski, 2015) çalışmalarında ortak filtreleme yöntemi için modifiye edilmiş DBSCAN algoritması sunulmuştur. Bir başka çalışmada,

DBSCAN algoritması öneri sistemlerindeki atak tespiti sırasında zamanı etki kullanmak için önerilmiştir (Das vd., 2014).

(Wu vd., 2012) çalışmalarında yarı-denetimli bir şilin atak tespit yöntemi sunulmuştur. Etkili tespit metriklerini seçmek için MC-Relief tanıtılmış ve yarı denetimli naive tabanlı bir sınıflandırıcı, normal kullanıcıları kötü niyetli kullanıcılardan ayırır. Deneysel değerlendirme için rastgele ve ortalama ataklar kullanılmıştır. Öneri sistemlerinde az sayıda etiketli kullanıcı vardır. Bu nedenle, (Cao vd., 2013)'de şilin ataklarını tespit etmek için yarı denetimli bir yöntem olarak naive tabanlı temel sınıflandırıcı önerilmiştir. EM-λ, önerilen sınıflandırıcının etkinliğini artırmak için kullanılmıştır. Son olarak, grafik tabanlı yarı denetimli bir algılama algoritması (Zhang vd., 2015) 'de sunulmuştur.

Öneri sistemlerinin yoğun kullanımı ile sistemlere yönelik atakların tespitine yönelik çalışmalar da artmıştır. Bu nedenle, bahsedilen gruplara dahil edilmeyen bazı araştırmalara da rastlamak mümkündür. Örneğin, (Wu vd., 2012)'da, naive tabanlı bir sınıflandırıcı etiketlenmiş veriler kullanılır, ardından etiketli verileri sınıflandırmak için bir sınıflandırma yaklaşımı kullanılır. Ayrıca kullanılan yöntem geliştirilmiş ve ürünlere ait bilgiler kullanılarak bir atak tespit yöntemi önerilmiştir (Wu vd., 2016).

Atak tespit yöntemleri için sınıflandırma ve kümeleme yöntemleri oldukça başarılı sonuçlar vermektedir. Yöntemler, normal kullanıcı sınıfına veya kümesine ait olmayan kullanıcıları tespit etmeyi amaçlar. Burada kullanıcılar nxm boyutlu uzayda bir nokta olarak değerlendirilir ve mesafeler dikkate alınarak bir sınıflandırma veya kümeleme yapılır. Bunlara ek olarak olasılık tabanlı yöntemlere de rastlamak mümkündür. Örneğin, Zhang vd.. bir olasılık yaklaşımı kullanarak sahte kullanıcıları tespit etmeye yönelik çalışma önermiştir (Zhang vd., 2006). Bhaumik vd. istatistiksel anormallik tespit yöntemini kullanarak atak tespit etmeye çalışmıştır (Bhaumik vd., 2006). Yöntemlerinde normal dağılım varsayımı vardır. X-bar ve güven aralığı kontrol limitleri kullanılarak standart sapma limitlerini aşan değerler belirlenmiştir.

2.3. Online Atak Tespit Yöntemleri

Ataklar, bilgisayar ve ağ sistemlerinin önde gelen problemleri arasında yer almaktadır. Bu nedenle bahsedilen sistemler ile ilgili çalışmaların çoğunluğu atak tespit yöntemlerini içermektedir. Bahsedilen ataklar bilgisayar sistemleri için bir anomali olarak

adlandırılabilir. Bilinen anormallik, ağ sistemlerine ciddi zararlar veren siber ataklardır (Harrou vd., 2019) . Bu atakları tespit etmek için farklı yöntemler önerilmektedir. Shewhart (Shewhart, 1926), KT (Page, 1954) ve ÜAHO (Roberts, 1959) olan kalite kontrol çizelgelerinin temel amacı, süreci kontrol etmek ve anormallikleri tespit etmek olsa da, bilgisayar ve ağ sistemlerinde anormallik tespiti için kullanılmaktadır. Bu grafikler, bilgisayar sistemlerindeki anormalliklerin tespiti için yüksek performans göstermiştir.

Hizmet Engelleme ve Dağıtılmış Hizmet Engelleme atakları, ağ sistemlerinin erişilebilirliğini ciddi şekilde etkilemektedir. Bu atakları tespit etmek için KT ve ÜAHO grafikleri kullanılmış ve yüksek tespit oranları elde edilmiştir (Bouyeddou vd., 2017). Ayrıca, (Yue vd., 2017) çalışmalarında ÜAHO çizelgeleri ile gerçek zamanlı bir anormallik tespit işlemi uygulanmıştır. Bahsedilen grafiklerin parametreleri algılama oranlarından etkilenebilir (Machaka vd., 2017). Önerilen çalışmada farklı parametre değerleri simüle edilmiş ve tespit oranları karşılaştırılmıştır. Başka bir çalışma (Ransewa vd., 2018) kaynak bağlantı noktası verilerine sistem dışı müdahalelere odaklanmıştır. Kalite kontrol grafiklerinin anormallik tespiti için farklı uygulamaları görmek mümkündür (Osanaiye vd., 2018; Sklavounos vd., 2017).

Şilin atakları, verilerdeki anormallikler olarak düşünülebilir. Öneri sistemi veri kümeleri zaman serisi verileridir ve şilin atakları gerçek kullanıcılardan farklılıklar sergilediği için şilin atakları bu veri kümesindeki anormalliklerdir. (Xia vd., 2015)'da, bu ataklar anormallikler olarak kabul edilir ve şilin ataklarını tespit etmek için dinamik bir zaman aralığı segmentasyonu uygulanmıştır. Ayrıca, atak profilleri veri kümesinin dağılımını değiştirmektedir. (Xu & Zhang, 2019)'de dinamik bir derecelendirme dağılım modeli önerilmiştir. Ardından, şilin ataklarını tespit etmek için derecelendirme kalıpları ve güven ilişkileri birleştirilerek DVM modeli oluşturulmuştur.

2.4. Shiryaev-Roberts Yöntemi

SR prosedürü, ortalama ve varyanstaki değişikliklerin hızlı bir şekilde tespit edilmesini sağlar. SR prosedürünün detayları (Shiryaev, 1961)'de sunulmuştur. İyi bilinen istatistiksel süreç kontrol araçları KT ve ÜAHO, ortalamadaki değişiklikleri tespit eder. Öte yandan, SR prosedürü, tek bir çizelgede ortalama ve varyans üzerindeki

değişikleri izler (Zhang vd., 2011). Ek olarak, SR prosedürü bilinmeyen ilk ortalama ve varyans ile ortalama değişiklikleri izleyebilir (Shauly-Aharonov vd., 2017).

Yüksek boyutlu verilerde gerçek zamanlı izleme için genel bir durma süresinin belirlenmesi için SR prosedüründe uyarlamalı örnekleme (W. Zhang & Mei, 2022)'de sunulmuştur. Ayrıca SR prosedüründe referans değerlerin büyük önemi vardır. (Lazariv & Schmid, 2018) 'de, değiştirilmiş SR prosedürü, sıralı olasılık oranı ve olabilirlik oranı değiştirilmiş SR prosedürü ile referans değerler olmadan uygulanmıştır.

(Polunchenko & Raghavan, 2018)'de KT ve SR çizelgeleri arasında karşılaştırmalı bir çalışma verilmiştir. Yöntemlerin yanlış alarm performansı ve ortalama algılama gecikmesi değerleri karşılaştırmalı olarak sunulmuştur. Küçük kaymaların tespiti için KT çizelgeleri önerilmiştir. KT çizelgeleri, küçük sapmalarda SR prosedüründen daha iyi performans gösterir. Öte yandan, SR prosedürü, durağan ortalama algılama gecikmesinde daha iyi performans gösterir (Moustakides vd., 2009). İstatiksel kalite kontrol araçları genellikle Gauss dağılımında iyi çalışılır. (Yu vd., 2022)'te ÜAHO, KT çizelgeleri ve SR prosedürü Weibull dağılımına uygulanmıştır. Çalışmanın sonuçları, $\beta \leq 1$ olduğunda SR prosedürünün KT ve ÜAHO çizelgelerinden üstün olduğunu göstermektedir. Benzer şekilde, SR prosedürü (Wu & Wu, 2022)'te Poisson süreçlerine uygulanmaktadır. Ancak bazı süreçlerde verilerin dağılımı bilinmeyebilir. (van Zyl & Lombard, 2022)'da dağılımdan bağımsız bir veriye SR prosedürü uygulanmıştır. Yakın tarihli bir çalışmada (Wu, 2020), KT ve SR, izole edilmiş değişen panellerin tahminini için melezleştirilmiştir.

2.5. Performans Ölçütleri

Mevcut literatürde, model doğrulama için performans ölçütleri önerilmektedir. Her metrik farklı amaçlar için kullanılmaktadır. Bazı metrikler pozitif değerlerin yanlış sınıflandırılmasını vurgularken, diğerleri negatif değerlerin yanlış sınıflandırılmasını vurgular. Tablo 2.3'de performans ölçütleri için karışıklık matrisi verilmiştir. Tablodan elde edilen değerler performans metrikleri hesaplanmasında kullanılmaktadır.

Tablo 2.3. *Performans ölçütleri için karışıklık matrisi*

		Gerçek Sınıf	
		Pozitif	Negatif
Tahmin Edilen Sınıf	Pozitif	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

Önerilen algoritmaların etkinliğini göstermek için bazı ölçütler kullanılmaktadır. İyi bilinen ve sıklıkla kullanılan metrik kesinlik (precision) oranıdır, diğer bir deyişle tespit oranıdır (Bhaumik vd., 2011; Gao vd., 2014; Hurley vd., 2009; Torres & Mendoza, 2019; Zhang, Chakrabarti, vd., 2006). Kesinlik oranı, Denklem 1.1'de gösterildiği gibi bir yanlış alarm oranı kullanılarak hesaplanır. Burada DP ve YP, sırasıyla doğru pozitif ve yanlış pozitif sayıları anlamına gelmektedir. Doğru-pozitif, doğru tespit edilen atak profillerini gösterir ve yanlış-pozitif, yanlış sınıflandırılmış, normal kullanıcı sayılarını gösterir.

$$Kesinlik = \frac{DP}{DP + YP} \tag{1.1}$$

Diğer en çok kullanılan değerlendirme metriği, duyarlılık (recall) oranıdır (Bhaumik vd., 2006; Mehta & Nejdl, 2009; Zapata vd., 2018) ve Denklem 1.2'de gösterildiği gibi hesaplanır; burada YN, yanlış sınıflandırılmış atak profillerinin sayısı gösterir.

$$Duyarlılık = \frac{DP}{DP + YN} \tag{1.2}$$

Bu çalışmada kullanılan son değerlendirme metriği F1 skorudur (Althbiti vd., 2021; Bhaumik vd., 2006; Burke vd., 2006b, 2006a). Bu metrik, kesinlik ve geri çağırma oranlarını birleştirir ve Denklem 1.3'teki gibi hesaplanır.

$$F1 - skoru = \frac{2\,x\,Duyarlılık\,x\,Kesinlik}{Duyarlılık + Kesinlik} \tag{1.3}$$

Doğruluk, doğru tahmin edilen gözlem sayısının tüm gözlem sayısına oranıdır. Doğruluk ölçütü Denklem 1.4'teki gibi hesaplanır. Sınıflandırma modelleri için yüksek doğruluk oranları tercih edilir. Ancak, sınıflar arasında sınıf dengesizliği sorunu olduğunda, model performansını sadece doğruluk oranına bakarak yorumlamak yeterli

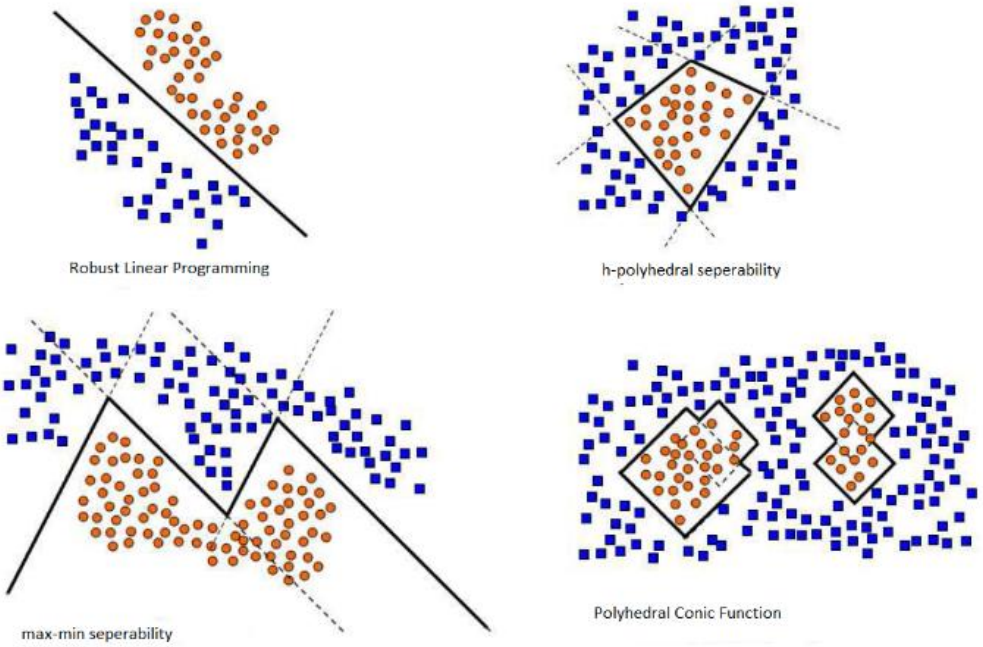
değildir. Pozitif gözlemlerin negatif gözlemlerden fazla olduğu durumlarda FP oranı doğruluk oranını çok fazla etkilemeyecektir. Bu nedenle, model performansını yorumlamak için yukarıda bahsedilen performans ölçütleri birlikte yorumlanır.

$$Doğruluk = \frac{DP + DN}{DP + DN + YP + YN} \tag{1.4}$$

2.6. Çokyüzlü Konik Fonksiyonlar ile Sınıflandırma ve Aşırı Örnekleme

Sınıflandırma algoritmaları hem akademik literatürde hem de gerçek dünyada geniş bir uygulama alanına sahiptir. Sınıflandırma algoritmalarının kullanıldığı bazı alanlar arasında beyin tümörünün tıbbi teşhisi (Kandimalla vd., 2022), metinlerin sınıflandırılması (Mirończuk & Protasiewicz, 2018), gezegen bilimi (Metzger vd., 2022) ve trafik modellemeleri (Atev vd., 2006) yer almaktadır. İyi bilinen sınıflandırma algoritmaları karar ağaçları (Barros vd., 2012), destek vektör makineleri ve yapay sinir ağlarıdır (Cervantes vd., 2020). Matematiksel programlama tabanlı yöntemler bir sonraki paragrafta ayrıntılı olarak verilmektedir.

Mangasarian ilk olarak Robust Linear Programming (RLP) (Bennett & Mangasarian, 1992) ile matematiksel programlama ile sınıflandırmayı başlatmıştır. A ve B, R^n'de boş olmayan sonlu kümeler olsun. Kümelerin dışbükey gövdelerinin kesişimleri boşsa, RLP kümeleri ayıran bir hiperdüzlem sağlar. Dışbükey gövdeler üzerinde bir kesişme varsa, RLP hiperdüzlem vererek hatayı en aza indirmeyi sağlar. h-çokyüzlü ayrılabilirlik yaklaşımı (Astorino & Gaudioso, 2002) 'da sunulmuştur. Bu yaklaşımda, kümelerden birinin dışbükey gövdesi diğer bir kümeyle kesişmiyorsa, h-çokyüzlü ayrılabilirlik bu kümeleri önceden belirlenmiş h-hiperdüzlemleri ile ayırabilir. h, hiperdüzlemlerin sayısıdır. Ayrıca, h-çokyüzlü ayrılabilirliğin genelleştirilmiş formu, Bagirov tarafından maksimum-minimum ayrılabilirlik olarak sunulmuştur (Bagirov, 2005). Kümeler, hiperdüzlemler ile oluşturulan bir hata fonksiyonunun en küçüklenmesiyle ayrılır. (Gasimov & Ozturk, 2006)'de, sınıflandırma problemlerinde konveks olmayan koniler elde etmek için Çok Yüzlü Konik Fonksiyonlar sunulmuştur. Bahsedilen dört matematiksel programlama tabanlı sınıflandırıcı, Görsel 2.1'de geometrik olarak gösterilmiştir.



Görsel 2.1. *Matematiksel sınıflandırıcıların geometrik gösterimi* (Gasimov & Ozturk, 2006)

2.7. Aykırı Değer Tespitine Dayalı Aşırı Örnekleme Tekniği

Bu çalışmada sınıflandırma için kullanılan yöntemlerden birisi çokyüzlü konik fonksiyonlardır. Ancak sınıf dengesizliği problemlerinde doğruluk oranları düşük olabilir. Bu dezavantajın üstesinden gelmek için azınlık sınıfı için aşırı örnekleme tekniği kullanılmıştır. En basit aşırı örnekleme tekniği, azınlık sınıfına ait örneklerin rasgele çarpıldığı rastgele aşırı örneklemedir (Random Over Sampling, ROS). (Kołcz akolcz vd., n.d.). Bu süreç aşırı uydurmaya neden olur. SMOTE yönteminde, k-en yakın komşu algoritması tarafından üretilen sentetik örneklerle bu sınırlamanın üstesinden gelinir (Chawla vd., 2011). Aykırı değer tespiti tabanlı yüksek hızda örnekleme tekniğinde (ODBOT) (Ibrahim, 2021), sınıflar arasındaki farklılık ilişkilerini göz önünde bulundurarak tutarlı sentetik örnekler sağlar. Ayrıca ODBOT çakışan problem riskini azaltır.

2.8. Destek Vektör Makineleri

Sınıflandırma algoritmaları, etiketli veriler ile daha önce görülmemiş verilerin sınıf etiketlerini tahmin eder. En iyi bilinen sınıflandırma algoritması olan DVM, sınıflandırma için matematiksel model kullanır. DVM'nin temel amacı eğitim örüntüleri ile karar sınırı arasındaki marjı maksimize etmektir (Boser vd., 1992). DVM, modellerin doğruluk oranlarını artırmak için çekirdekler kullanır.

DVM, akademik literatürde ve gerçek dünyada geniş bir uygulama alanına sahiptir. (Alostad, 2019)'de, Gauss Karışım Modeli ile şilin saldırı tespiti için modifiye edilmiş bir DVM modeli kullanılmıştır. Benzer şekilde, DVM ve Hedef Ürün Analizi şilin saldırısı tespitinde kullanılmıştır. Ayrıca, SMOTE sınıf dengesizliği probleminin üstesinden gelmek için kullanılmıştır (Zhou vd., 2016).

3. YÖNTEM

Bu tez çalışmasında şilin ataklarının çevrimiçi tespiti için Kümülatif Toplam (KT), Üstel Ağırlıklandırılmış Hareketli Ortalama (ÜAHO) ve Shiryaev-Roberts kalite kontrol grafikleri kullanılmıştır. Devam eden bölümlerde ilgili grafiklerin matematiksel yöntemleri verilmiştir. Kalite kontrol grafikleri ile tespit edilen şüpheli kullanıcılar arasından atak kullanıcıları tespit etmek için korelasyon matrisi üzerinden kümeleme yapan yoğunluk tabanlı kümeleme algoritması kullanılmıştır. Öneri sisteminin gürbüzlüğünün devamı için tespit edilen atak kullanıcılar çok yüzlü konik fonksiyon sınıflandırıcısı ile eğitilerek gelecekteki benzer atakların önüne geçilmiştir. Atak kullanıcılar, geçek kullanıcılara göre sayı olarak daha az olduğu için aşırı örnekleme yöntemi ile doğruluk oranının artırılması hedeflenmiştir. Devam eden kısımda ilgili yöntemlerin matematiksel formülleri ve açıklamaları verilmektedir.

3.1. Kalite Kontrol Grafikleri

Verilerin zaman içindeki çeşitliliğini araştırmak için kalite kontrol grafikleri önerilmektedir. Bu çeşitlilik, insan hatalarından, operasyonel değişikliklerden ve sürecin doğasından kaynaklanmaktadır. Kalite kontrol çizelgeleri, grafikleri çeşitliliğin kaynağını araştırır. Bu çalışmada literatürde sıklıkla kullanılan KT, ÜAHO ve SR kalite kontrol grafikleri kullanılmıştır.

3.1.1. Kümülatif Toplam (KT)

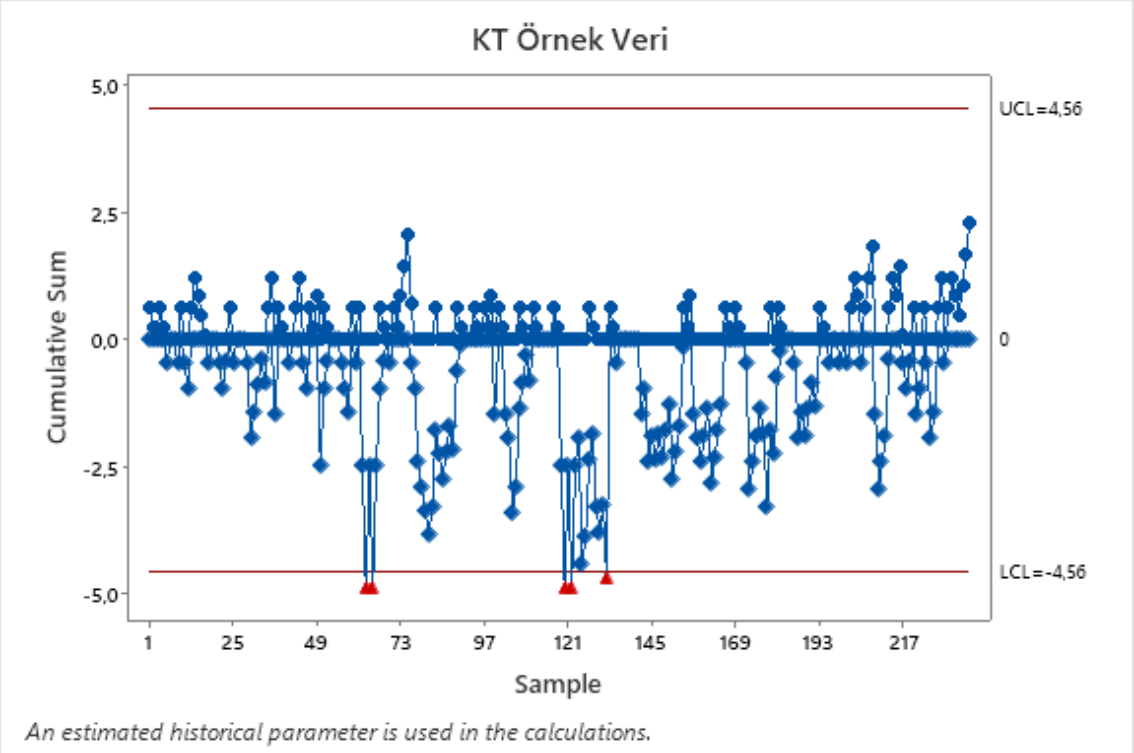
KT grafikleri, ortalama değerden küçük sapmaları tespit etmek için Page (1954) tarafından önerilmiştir.

Şekil 3.1'de KT grafiğinin MovieLens 100K veri setinde bir profil üzerine yapılış puanlamaların uygulaması sunulmuştur. Burada kırmızı üçgen ile tespit edilen noktalar normal verinin haricinde dışarıdan eklenen verilerdir. KT istatistiği alt ve üst kontrol limitlerini aştığı durumda şekilde olduğu gibi sinyal vermektedir.

Ortalama değerlerden sapmalar kümülatif olarak toplanır ve grafikte işaretlenir. KT istatistiği Denklem 3.1'deki gibi hesaplanır.

$$C_i = \sum_{i=1}^{i} \left(x_i - \mu_0 \right) \tag{3.1}$$

Burada C_i ortalamadan i. değere kadar olan sapmaların kümülatif toplamı, x_i ise i. Gözlemin değeridir. μ_0 ortalama veya hedeflenen değerdir. KT grafikleri, veriler göz önüne alındığında farklı eğilimler göstermektedir. İşlem kontrol altındaysa, C_i sıfır toplamlı bir dağılım sergiler. Ancak, işlem ortalaması, μ_1, μ_0'dan büyükse, C_i'de pozitif bir eğilim izlenir. Öte yandan, μ_1, μ_0'dan küçükse, C_i'de negatif bir eğilim izlenir. Pozitif ve negatif eğilimler sırasıyla Denklem 3.2 ve 3.3'te gösterilmiştir.



Şekil 3.1. *KT grafiği örnek veri üzerine uygulama*

$$C_i^+ = \max\left[0, x_i - \left(\mu_0 + K\right) + C_{i-1}^+\right] \tag{3.2}$$

$$C_i^- = \max\left[0, \left(\mu_0 - K\right) - x_i + C_{i-1}^+\right] \tag{3.3}$$

C_i^+ ve C_i^-'nin başlangıç değerlerinin sıfırdır. K tespit edilmek istenen sapma miktarıdır ve Denklem 3.4-3.6'da hesaplanması verilmiştir.

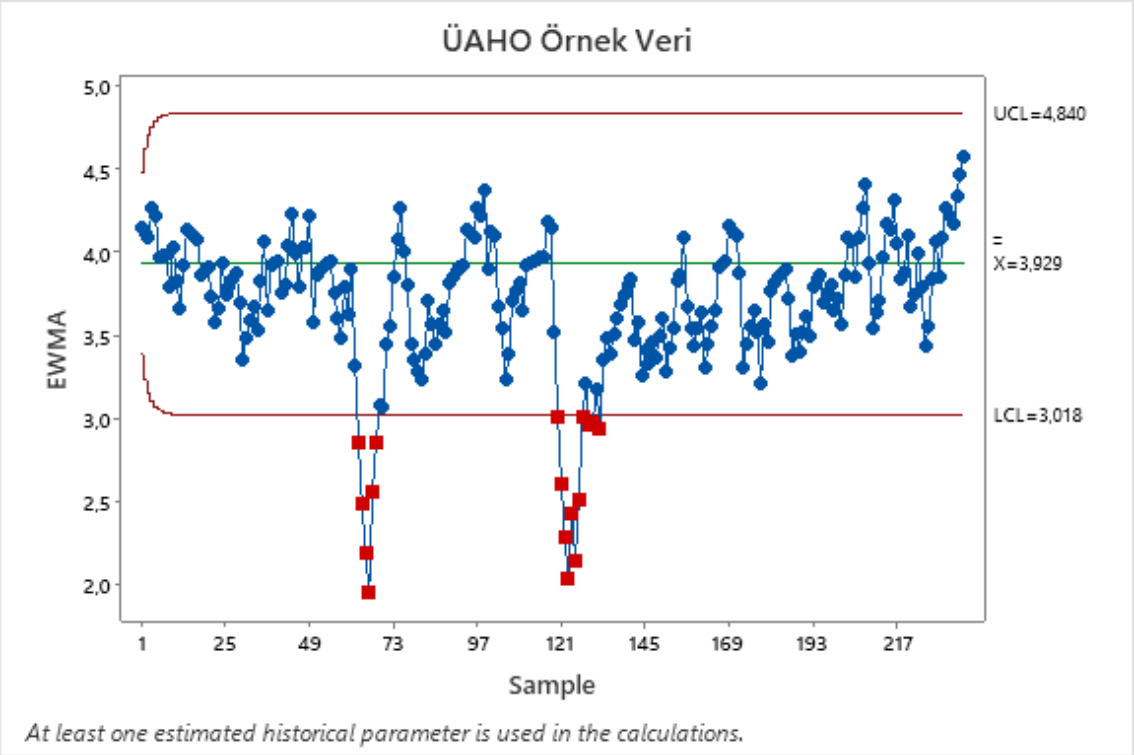
$$\mu_1 = \mu_0 + \delta\sigma \tag{3.4}$$

$$\delta = \frac{\left|\mu_1 - \mu_0\right|}{\sigma} \tag{3.5}$$

$$K = \frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_0|}{2} \tag{3.6}$$

3.1.2. Üstel Ağırlıklandırılmış Hareketli Ortalama (ÜAHO)

ÜAHO grafiği KT grafiğine bir alternatif olarak Roberts (Roberts, 1959) tarafından önerilmiştir. ÜAHO yaklaşımı ağırlıklandırılmış gözlemlere dayanmaktadır ve bu ağırlıklandırma geçmişe doğru üstel azalmaktadır. KT grafiğinde olduğu gibi hesaplanan istatistik limitleri aştığı takdirde grafik Şekil 3.2'de olduğu gibi sinyal vermektedir.



Şekil 3.2. *ÜAHO grafiği örnek veri üzerine uygulama*

ÜAHO istatistiği Denklem 3.7'de verilmiştir.

$$Z_i = \lambda x_i + (1-\lambda) Z_{i-1} \tag{3.7}$$

Burada x_i i. gözlem değeri, Z_i i. gözlem değerinin de içinde olduğu ağırlıklandırılmış istatistiktir. λ ağırlıklandırma parametresidir ve 0<λ≤1 aralığında değer. λ değeri azaldıkça geçmiş puanların mevcut istatistiğe etkisi artacaktır. Grafiğin üst ve alt sınırları sırasıyla Denklem 3.8 ve 3.9'da verilmiştir.

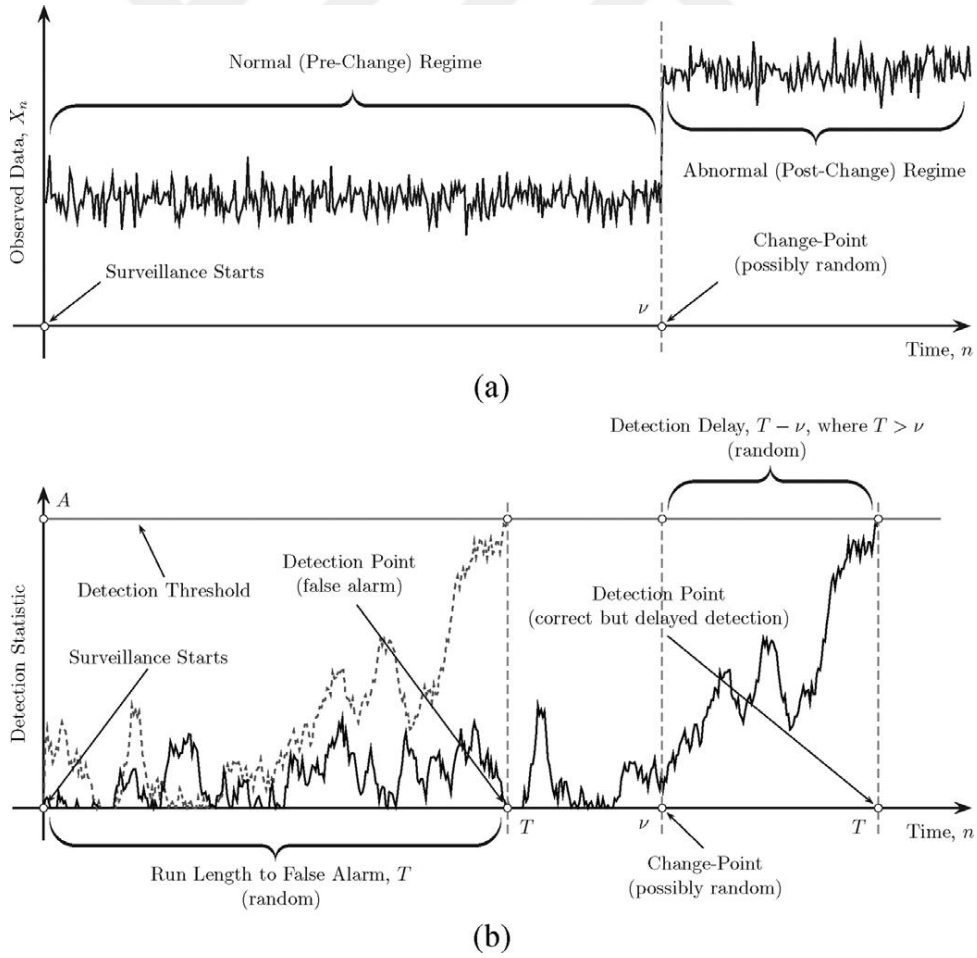
$$\mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{3.8}$$

$$\mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]} \tag{3.9}$$

Burada μ_0 mevcut işlemin ortalama değeridir ve L, ÜAHO'nun sabit parametresidir. Farklı L değerleri gözlemlenebilir. L ve λ değerlerinin farklı kombinasyonları, grafiğin hassasiyetini azaltır veya artırır. KT çizelgelerinden farklı olarak, her gözlem için kontrol limitleri farklıdır.

3.1.3. Shiryaev-Roberts (SR) Prosedürü

SR grafiği aynı anda ortalama ve varyans üzerindeki değişikliği takip etmeye yarayan kalite kontrol grafiğidir. Şekil 3.3 (a)'da değişim noktası, v'den önceki kısım normal süreci gösterirken değişim noktasından sonraki sürecin görseli verilmiştir. Şekil 3.3 (b)'de ise yanlış alarm tasvir edilmiştir. Değişim noktasından önce grafik istatistiğinin yanlışlıkla alarm verdiği gösterilmiştir.



Şekil 3.3. *Shiryaev-Roberts grafiği görseli* (Tartakovsky vd., 2013)

$x_t = (x_{t1}, \dots, x_{tn})$ örneklem boyutu $n \geq 1$ olan ve x verisinin kalite karakteristiklerini taşıyan bir örneklem olsun. Bu örneklemin ortalama ve standart sapma üzerinde bir sapma gerçekleşirse, sürecin ortalaması Denklem 3.10'daki gibi olmaktadır.

$$x_{ti} = \mu + \varepsilon_{ti}, \quad \begin{matrix} i = 1,...,n, \\ t = 1,2,... \end{matrix} \tag{3.10}$$

Burada $\varepsilon_{t1}, \dots, \varepsilon_{ti}$ ortalaması 0 standart sapması σ olan bağımsız ve özdeşçe dağılmış normal dağılımdan gelmektedir. Kontrol altındaki durumda ortalama ve standart sapma sırasıyla $\mu = \mu_0$, $\sigma = \sigma_0$ olarak tanımlanır. Eğer süreç kontrol dışına çıkarsa yeni ortalama ve standart sapma $\mu_1 = \mu_0 + \delta\sigma_0$ ve $\sigma_1 = \gamma\sigma_0$, olacaktır.

$t \geq 1$, için SR istatistiği Denklem 3.11'deki gibi hesaplanmaktadır. Burada $\Lambda_t = g(x_t) \Big/ f(x_t)$ olarak tanımlanır ve $f(x_t)$ kaymadan önceki dağılımı gösterirken, $g(x_t)$ kaymadan sonraki dağılımı göstermektedir.

$$\Lambda_t = \gamma^{-1} \exp\left\{ \frac{\gamma^2 - 1}{2\gamma^2} \sum_{j=1}^{n} x_{tj}^2 + \frac{\delta}{\gamma^2} \sum_{j=1}^{n} x_{tj} - \frac{n\delta^2}{2\gamma^2} \right\} \tag{3.11}$$

SR prosedürünün durdurma kriteri Denklem 3.12'da verilmiştir.

$$R_t > h \tag{3.12}$$

Durdurma kriteri R_t Denklem 3.13'te gösterildiği gibi hesaplanmaktadır.

$$R_t = \sum_{k=1}^{t} \prod_{j=k}^{t} \Lambda_t \tag{3.13}$$

Durdurma kriteri R_t Denklem 3.14'te gösterildiği gibi özetlenmektedir.

$$R_t = \left(1 + R_{t-1}\right) \Lambda_t \tag{3.14}$$

3.2. Korelasyon Analizi

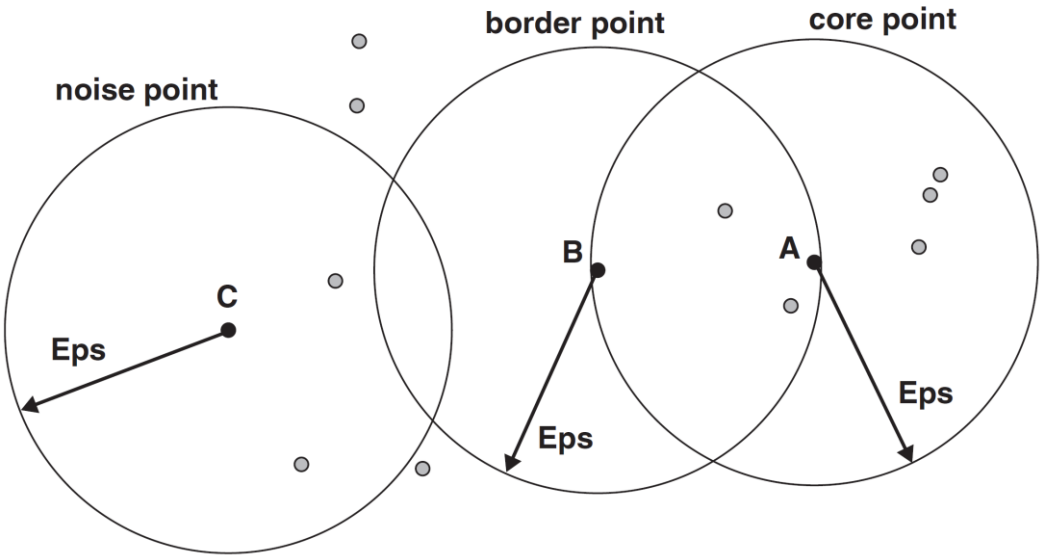
Korelasyon iki rasgele değişken arasındaki istatiksel benzerlik olarak tanımlanmaktadır. Bu çalışmada korelasyon iki kullanıcı arasındaki ilişkiyi ölçmek için kullanılmış ve Denklem 3.15'teki gibi hesaplanmıştır.

$$\rho_{X,Y} = \frac{E\left[\left(X-\mu_x\right)\left(Y-\mu_y\right)\right]}{\sigma_X \sigma_Y} \tag{3.15}$$

Burada $\rho_{X,Y}$ iki rastgele değişken X ve Y arasındaki popülasyon korelasyon katsayısıdır. μ_x ve μ_y, sırasıyla X ve Y'nin beklenen değerleridir. E, verilen denklemin beklenen değeri ve σ_X ve σ_Y'nin X ve Y'nin standart sapmasını göstermektedir.

3.3. DBSCAN Algoritması

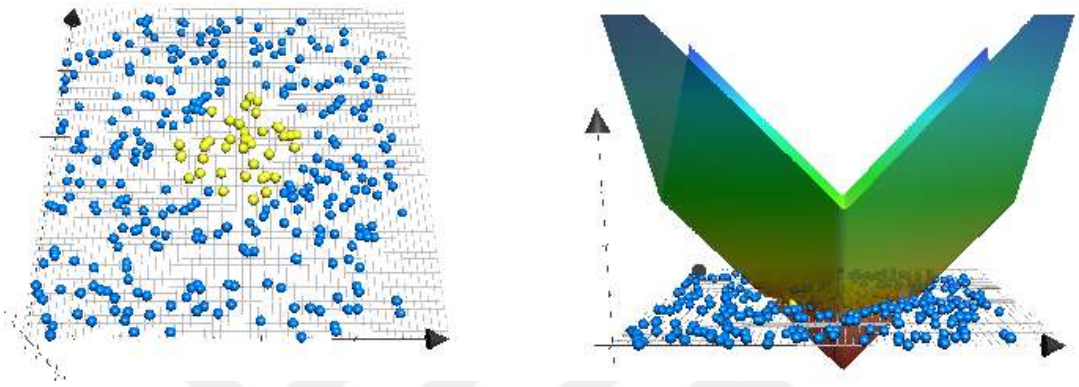
Rastgele şekillerin ve gürültü verilerin tahmini için önerilen yoğunluk tabanlı DBSCAN kümeleme algoritması epsilon (Eps) ve minimum nokta (MinPts) olmak üzere iki parametreye sahiptir. En basit tanımla, algoritma epsilon komşuluğunda minimum nokta içeren kümenin oluşturulması ile kümeleme işlemine başlamaktadır. Oluşturulan kümenin içindeki noktaların aynı yöntemle komşuluk araması ile genişletilmektedir. DBSCAN algoritmasında noktalar üç başlık altında incelenmektedir. Çekirdek nokta, epsilon çapındaki komşuluğunda en az minimum nokta kadar komşuya sahip olan noktalardır. Sınır noktaları ise epsilon komşuluğunda minimum noktadan az nokta içeren fakat bir çekirdek noktanın komşuluğunda olan noktalardır. Son olarak, gürültü noktaları ise epsilon komşuluğunda minimum noktadan az nokta olan ve herhangi bir çekirdek noktanın komşuluğunda olmayan noktalardır. Bahsedilen noktaların görsel hali Görsel 3.1'de verilmiştir.



Şekil 3.4. *DBSCAN algoritması parametreleri ve noktaları*

3.4. Çokyüzlü Konik Fonksiyonlar

Bu kısımda çokyüzlü konik fonksiyonların metodolojisi özet olarak verilmiştir. ÇKF'de sınıflandırma işlemi çokyüzlü konik fonksiyonlar tarafından üretilen çokyüzlü konik bölgeler ile sağlanmaktadır. Görsel 3.1'de CKF'nin ayırma mekanizması görsel olarak sunulmuştur. Görselin sol tarafında rasgele üretilmiş veriler bulunmaktadır. Sağ tarafta ise bu verilerin CKF ile ayrımı gösterilmektedir.



Görsel 3.1. *Çokyüzlü konik fonksiyonların ayırma görseli* (Cevikalp vd., n.d.)

ÇKF ile ilgili çalışmalar (Gasimov & Ozturk, 2006; Ozturk & Ciftci, 2014)'de sunulmuştur. Çokyüzlü konik fonksiyon Denklem 3.16'da verilmiştir.

$$g_{(w,\xi,\gamma,c)} = \langle w,(x-c)\rangle + \xi\|x-c\|_1 - \gamma \tag{3.16}$$

Burada $w \in R^n$, öznitelikler için ağırlık fonksiyonu, $c \in R^n$ kümelerin merkez noktasını, $\xi \in R^+$ ağırlık katsayısı ve $\gamma \in R^+$ise offset değer olarak tanımlanmaktadır. $x \in R^n$ ise sınıfı belirlenecek gözlemdir. ÇKF için kullanılacak iki sınıf Denklem 3.17 ve 3.18'de tanımlanmıştır.

$$A = \left\{a^1,...,a^m\right\}, a^i = \mathbb{R}^n, i \in I = \{1,...,m\} \tag{3.17}$$

$$B = \left\{b^1,...,b^p\right\}, b^j = \mathbb{R}^n, j \in J = \{1,...,n\} \tag{3.18}$$

Matematiksel programlama algoritmalarında pozitif gözlemler için pozitif değerler elde edilir. Ancak ÇKF'lerde pozitif değerler için negatif ÇKF değerleri elde edilir.

3.5. Aykırı Değer Tespitine Dayalı Aşırı Örnekleme Tekniği

Sınıflandırma probleminde sınıf dengesizliği probleminin üstesinden gelmek için aykırı değer tespitine dayalı aşırı örnekleme tekniği önerilmiştir. ODBOT algoritması ayrıntılı olarak (Ibrahim, 2021)'te açıklanmıştır. Algoritma adımları aşağıda verilmiştir.

Adım 1. *Sentetik numune sayısının (NSS) belirlenmesi:* Gerekli sentetik numune sayısı, yüzde değeri (PV) ile azınlık sınıfındaki numune sayısının çarpılmasıyla elde edilir. Gerekli sentetik numune sayısı Denklem 3.19 ile hesaplanmaktadır.

$$NSS = \frac{PV * |minority\ class|}{100} \tag{3.19}$$

Buradaki yüzde değeri Denklem 3.20 ile hesaplanmaktadır.

$$PV = \left(\left(\frac{No.of\ samp.in\ MajClu}{No.of\ samp.in\ MinClu} \right) - 1 \right) * 100 \tag{3.20}$$

Adım 2. *Farklılık ilişkilerinin bulunması:* sınıflar arasındaki farklılığın sınıflandırmada önemli bir rolü vardır. Bu nedenle ODBOT algoritması, sınıf farklılıklarını belirlemek için k-ortalamalar algoritması ile melezlenmiş ağırlık tabanlı bir yarasa algoritması kullanır.

Adım 3. *Aykırı değerlerin tespiti:* Aykırı değerler sınıflandırma algoritmalarına zarar verir ve sınıflandırma doğruluğunu olumsuz etkiler. ODBOT algoritması, Denklem 3.21'de tanımlanan azınlık küme mesafesinin toplamını hesaplayarak aykırı değerleri tespit eder.

$$SMCD_{i,j} = \sum_{l=1}^{k} \left| MinCen_{i,j} - MajCen_{l} \right| \tag{3.21}$$

En iyi azınlık sınıfı Denklem 3.22 ile seçilmektedir.

$$Max\left\{ SMCD_{i,1}, SMCD_{i,2}, ..., SMCD_{i,k} \right\} \tag{3.22}$$

Adım 4. *Sentetik numunelerin oluşturulması:* Sentetik numuneler, en iyi azınlık sınıfları sınırı dikkate alınarak üretilir. Sentetik numunelerin üretilmesi Denklem 3.23'te gösterilmiştir.

$$S = r * (Max\_Min - Min\_Min) + Min\_Min \tag{3.23}$$

Burada S, sentetik numuneleri ve r, 0 ile 1 arasındaki rastgele sayıdır. Max_Min ve Min_Min sırasıyla en iyi azınlık sınıfındaki maksimum ve minimum değerlerdir.

3.6. Destek Vektör Makineleri

Destek vektör makineleri iki sınıfın verilerini ayırabilecek en iyi ayırma hiperdüzlemini vermektedir. Burada kullanılacak öğrenme kümesi $D = \left\{ (x_i, y_i) \,|\, i = 1...N, x_i \in R^d \; ve \; y_i \in \{-1, +1\} \right\}$ olsun. Pozitif örneklemlerin bulunduğu sınıf C_1, negatif örneklemlerin bulunduğu sınıf C_2 olsun.

Destek vektör makinesi matematiksel modeli Denklem 3.24'de verilmiştir.

$$\begin{aligned} &\max_{D,w,w_0} \frac{2D}{\|w\|} \\ &s.t. \quad \left(w^T x_i + w_0\right) \geq D, \quad \forall x_i \in C_1 \\ &\qquad \left(w^T x_i + w_0\right) \leq -D, \quad \forall x_i \in C_2 \end{aligned} \tag{3.24}$$

Burada amaç fonksiyonu ile marjlar arası mesafe en büyüklenmek istenmektedir. İlk kısıt pozitif sınıftaki elemanlar, ikinci kısıt ise negatif sınıftaki elemanlar içindir.

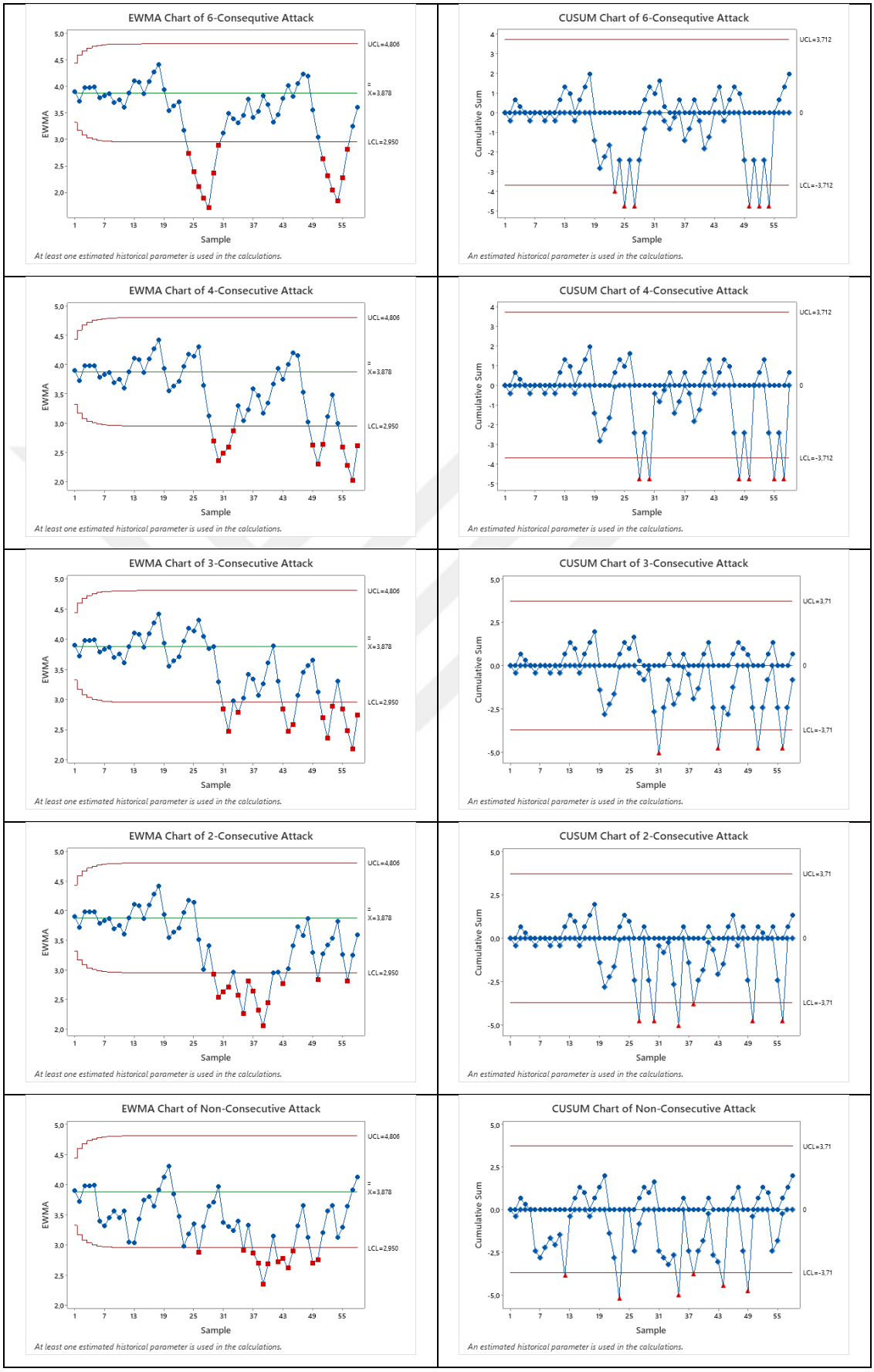
4. DENEYSEL SONUÇLAR

Bu çalışmada, deneysel çalışmalar için MovieLens 100K ve 1M veri setleri kullanılmıştır. 100K[1] veri seti 1682 ürüne 943 kullanıcı tarafından verilmiş 100.000 puan bulunmaktadır. Her kullanıcı en az 20 filme puan vermektedir. Bu veri kümesi yüksek seyrekliğe sahiptir ve seyreklik oranı yaklaşık %93,7'dir. MovieLens100K veri kümesi 5 yıldızlı derecelendirmeye sahiptir. Puanlar {1,2,3,4,5} kümesinde değerler alır. Veri kümesi dört sütunlu veriler içerir. Kullanıcı ve film kimlikleri sırasıyla birinci ve ikinci sütunlarda verilmiştir. Verilen puanlamalar üçüncü sütunda yer almaktadır. Verilen puanların zaman damgaları dördüncü sütuna yerleştirilmiştir.1M[2] veri seti ise 3952 ürüne 6040 kullanıcı tarafından yapılmış 1.000.209 puanlamadan oluşmaktadır.

4.1. Şüpheli Puanlamaların Kalite Kontrol Grafikleri ile Tespiti

Sistemlerde anomaliler farklı örüntüler halinde görülebilir. Anomalilerin çoğu ardışık bir örüntü gösterir. Benzer şekilde, şilin atakları, atak sayısı arttıkça ardışık bir model izler, tespit oranı artar. Aşağıdaki şekillerde görüldüğü gibi aynı veri seti için 6, 4, 3, 2 ve ardışık olmayan ataklar incelenmiştir. Her durumda, sonuç dikkate alınarak veri kümesine rastgele 12 atak eklenmiştir. KT ve ÜAHO grafikleri tüm ardışık durumlarda atakları tespit etmiştir. Şekil 4.1'de kullanılan ÜAHO grafiğinde ataklardan sonra istatistik sıfırlanmamıştır, bu nedenle bazı atak dışı (gerçek) puanlamalar atak olarak tespit edilmiştir. Bu sorun, atakların komşularının ayrıntılı olarak araştırılmasıyla çözülmüştür. KT grafiğinde ise atak puandan sonra istatistik sıfırlanmıştır. Bu, bir atak meydana geldiğinde, KT grafiğinin durdurulduğu ve ardından başlangıç değerleriyle başlatıldığı anlamına gelir. Bu durumda, bazı atak puanları normal puanlamalar gibi görünmektedir. ÜAHO şemasında olduğu gibi, atakların komşuları ayrıntılı olarak araştırılmaktadır. Böylece yanlış-pozitif ve doğru-negatif oranlarının azaltılması planlanmıştır.

[1] https://grouplens.org/datasets/movielens/100k/
[2] https://grouplens.org/datasets/movielens/1m/



Şekil 4.1. *Farklı ardışık atakların veriye etkisi*

4.1.1. ÜAHO için parametre seçimi

ÜAHO grafikleri için iki parametre vardır, L ve lambda. L alt ve üst limitlerin belirlenmesinde kullanılır. Daha yüksek L, düşük yanlış alarm oranları sebep olan daha geniş bir aralık anlamına gelir. Ancak daha yüksek bir L değeri, şüpheli oranın tespit oranının düşmesine neden olur. Aynı zamanda L sayısının küçük olması, yüksek algılama oranıyla daha küçük bir aralık anlamına gelir.

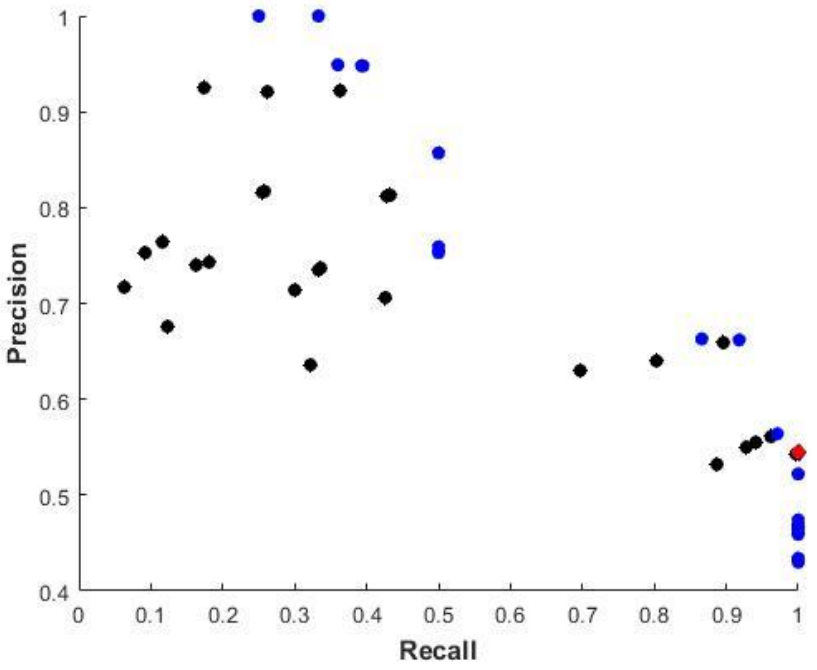
Bununla birlikte, düşük L değerleri yüksek yanlış alarm oranlarına neden olur. Diğer bir parametre lambda, geçmiş veriler için unutma oranıdır. Lambda değeri yüksekse, geçmiş verilerin etkisi küçüktür, bunun tersi de geçerlidir. Her iki parametre de geri çağırma ve kesinlik oranları için gereklidir.

Optimal parametre düzeyine karar vermek için ardışık tüm tipler için 100 test verisi kullanılmıştır. Her bir test verisine 46 gerçek puanlama ve 12 atak puanlama. Ek olarak, atak puanlamalar verileri test etmek için rastgele yerleştirilmiştir. Bu veri seti, optimal parametre seviyelerine karar vermek için her iki kontrol grafiği için de kullanılmıştır. Her iki parametre için beş farklı seviye tanımlanmıştır ve ÜAHO grafiği için Tablo 4.1'de farklı hatırlama ve kesinlik sonuçları kombinasyonları verilmiştir. Tablolardan da görülebileceği gibi, L=1,5 ve λ =0,8, yüksek algılama oranı için en uygun değerlerdir.

Bu problemde yüksek hassasiyet ve hatırlama oranı hedeflenmektedir. Ancak, bahsedilen metrikler çelişkilidir. Bu çelişki Görsel 4.2'de görülebilir. Şekilde siyah noktalar Pareto olmayan çözümleri ifade etmektedir. Mavi noktalar ve kırmızı nokta Pareto çözümleri göstermektedir. Bu çözümler herhangi başka bir nokta tarafından baskılanmamaktadır. Her iki ölçüyü de maksimize eden tek bir çözüm yoktur. Bu nedenle ağırlıklı toplam skalarizasyon yöntemi kullanılarak bir Pareto optimal (kırmızı nokta) çözüm elde edilmiştir (Kasimbeyli et al.vd., 2017)[85]. Geri çağırma metriği, kesinlik metriğine göre daha önemlidir. Bu nedenle, Tablo 4.1'de vurgulanan değer, Pareto optimal çözüm olarak seçilmiştir. L = 1.5 ve λ = 0.8, bu problem için seçilen Pareto optimal değerdir.

Tablo 4.1. *ÜAHO parametre seçimi*

		Duyarlılık					Kesinlik				
	L/λ	0,5	0,6	0,7	0,8	0,9	0,5	0,6	0,7	0,8	0,9
1	1	1	1	1	1	1	0,474	0,474	0,434	0,459	0,459
	1,50	0,697	0,887	0,997	1	1	0,630	0,532	0,543	0,544	0,522
	2	0,322	0,336	0,336	0,333	0,300	0,636	0,737	0,737	0,735	0,714
	2,50	0,181	0,181	0,123	0	0	0,743	0,743	0,676	NaN	NaN
	3	0,063	0	0	0	0	0,717	NaN	NaN	NaN	NaN
2	1	1	1	1	1	1	0,469	0,469	0,431	0,460	0,460
	1,50	0,803	0,928	1	1	1	0,640	0,550	0,544	0,544	0,522
	2	0,500	0,500	0,500	0,500	0,500	0,759	0,857	0,857	0,857	0,857
	2,50	0,258	0,255	0,163	0	0	0,817	0,816	0,740	NaN	NaN
	3	0,092	0	0	0	0	0,753	NaN	NaN	NaN	NaN
3	1	1	1	1	1	1	0,466	0,466	0,430	0,460	0,460
	1,50	0,866	0,941	1	1	1	0,663	0,555	0,545	0,545	0,522
	2	0,426	0,433	0,433	0,432	0,428	0,706	0,813	0,813	0,813	0,812
	2,50	0,360	0,360	0,333	0	0	0,949	0,949	1,000	NaN	NaN
	3	0,116	0	0	0	0	0,764	NaN	NaN	NaN	NaN
4	1	1	1	1	1	1	0,464	0,464	0,430	0,461	0,461
	1,50	0,896	0,962	1,000	1	1	0,659	0,561	0,545	0,545	0,522
	2	0,500	0,500	0,500	0,500	0,500	0,755	0,857	0,857	0,857	0,857
	2,50	0,363	0,363	0,262	0	0	0,922	0,922	0,921	NaN	NaN
	3	0,250	0	0	0	0	1,000	NaN	NaN	NaN	NaN
6	1	1	1	1	1	1	0,464	0,464	0,430	0,461	0,461
	1,50	0,918	0,971	1	1	1	0,662	0,564	0,545	0,545	0,522
	2	0,500	0,500	0,500	0,500	0,500	0,753	0,857	0,857	0,857	0,857
	2,50	0,395	0,393	0,333	0	0	0,948	0,948	1,000	NaN	NaN
	3	0,174	0	0	0	0	0,925	NaN	NaN	NaN	NaN



Görsel 4.1. *Çelişen amaçların görseli*

4.1.2. KT için parametre seçimi

İki parametre, K ve h, KT grafiğinin tespit performansını etkiler. K, tespit edilecek sapmayı gösterir ve h, grafiğin sınırıdır. Kümülatif değer h değerini aşarsa, grafik tespit sinyali verecektir. Tablo 4.2'de, ardışık atakların farklı kombinasyonlarının geri çağırma ve kesinlik değerleri, K ve h değerlerinin farklı kombinasyonları için verilmiştir. Bahse konu veri seti için ÜAHO grafiklerinin daha iyi sonuçlar verdiği gözlemlenmiştir.

Hem ÜAHO hem de KT tablolarının sonuçları göz önüne alındığında, ÜAHO grafiği bazı kombinasyonlar için yüksek oranlar vermektedir. Ek olarak, yüksek hassasiyetli metrik, atak olmayan derecelendirmelerin atak derecelendirmesi gibi olduğu anlamına gelir. Bahsedilen sorun daha sonraki adımlarda çözülebilir. Bu şüpheli derecelendirmeler, kümeleme adımında araştırılır. Öte yandan, düşük bir geri çağırma oranı ciddi bir sorundur, çünkü bir atak derecesi atak olmayan bir derecelendirme olarak etiketlenirse, bu sorun kümeleme adımında çözülemez. Bahsedilen sorunlar göz önüne alındığında bu çalışmada yüksek hatırlama oranları istenmektedir. Bu nedenle bu çalışmada atak tespiti için bir ÜAHO şeması önerilmiştir.

Tablo 4.2. *KT parametre tahmini*

		Duyarlılık					Kesinlik				
	K/h	2,0	3,0	4,0	5,0	6,0	2,0	3,0	4,0	5,0	6,0
1	0,25	0,420	0,344	0,138	0,032	0	0,553	0,614	0,692	0,844	NaN
	0,50	0,328	0,158	0,064	0,002	0	0,543	0,686	0,664	1	NaN
	1	0,188	0,140	0,035	0	0	0,676	0,683	0,627	NaN	NaN
	1,50	0,153	0,084	0,013	0	0	0,694	0,664	0,455	NaN	NaN
	2	0,097	0,058	0,004	0	0	0,663	0,676	0,714	NaN	NaN
2	0,25	0,500	0,500	0,158	0,038	0	0,608	0,736	0,683	0,836	NaN
	0,50	0,500	0,212	0,103	0,007	0	0,663	0,730	0,783	0,727	NaN
	1	0,218	0,168	0,048	0,002	0	0,685	0,698	0,722	1	NaN
	1,50	0,183	0,133	0,025	0,001	0	0,724	0,766	0,714	1	NaN
	2	0,142	0,075	0,011	0,000	0	0,742	0,776	0,722	NaN	NaN
3	0,25	0,456	0,433	0,333	0,053	0	0,556	0,675	1	0,750	NaN
	0,50	0,417	0,333	0,103	0,026	0	0,589	1	0,769	0,939	NaN
	1	0,333	0,333	0,067	0,008	0	0,941	1	0,833	1	NaN
	1,50	0,333	0,115	0,053	0	0	1	0,687	0,928	NaN	NaN
	2	0,129	0,090	0,029	0	0	0,701	0,715	0,946	NaN	NaN
4	0,25	0,500	0,500	0,271	0,060	0	0,586	0,713	0,923	0,661	NaN
	0,50	0,500	0,272	0,250	0,028	0	0,636	0,842	1,000	0,971	NaN
	1	0,272	0,271	0,074	0,012	0	0,753	0,923	0,712	0,778	NaN
	1,50	0,272	0,250	0,045	0,003	0	0,924	1	0,771	1	NaN
	2	0,250	0,118	0,033	0,001	0	1,000	0,793	0,889	0,500	NaN
6	0,25	0,500	0,500	0,333	0,167	1	0,579	0,703	1,000	1	NaN
	0,50	0,500	0,333	0,176	0,034	1	0,624	1	0,934	0,612	NaN
	1	0,333	0,333	0,167	0,007	0,500	0,887	1	1	1	NaN
	1,50	0,333	0,183	0,167	0,007	0	1	0,852	1	1	NaN
	2	0,183	0,167	0,038	0,003	0	0,849	1	0,648	1	NaN

4.2. Korelasyon Katsayısının Uzaklık Olarak Kullanılması

Literatürde farklı ihtiyaçlar doğrultusunda farklı mesafe ölçütleri geliştirilmiştir. Manhattan, Öklid, minkowski, jaccard ve korelasyon bu metriklerin başında gelmektedir. Her metrik aynı noktalar arasındaki mesafeyi farklı uzunluklarda ifade edebilmektedir. Bu çalışmada kümeleme işleminin gerçekleştirilmesi için uzunluk ölçütü olarak korelasyon kullanılmıştır. Öneri sistemlerinde kullanıcıların puanlayabileceği çok fazla ürün bulunmaktadır. Bu durumda, diğer metrikler için çok geniş aralıklarda ifade edilen uzaklık birimleri, korelasyon ile -1 ile +1 aralığında ifade edilebilmektedir. Ayrıca korelasyon yakınlığın yanında birlikte iki kullanıcın arasındaki benzerlik arasında da bilgi vermektedir. Korelasyon katsayılarının uzaklık olarak kullanılması ile birlikte kümeleme için seçilecek yöntemin parametre seçimi daha dar aralıkta yapılacağı için parametre seçimi daha kolay yapılacaktır. Tablo 4.3'te Movilens100K veri setinde rastgele bir ürüne yapışmış %1 atak ve doldurma oranına sahip rastgele atağın gerçek kullanıcılar arasındaki hesaplanmış korelasyonu verilmiştir. Bu çalışmada, literatürün aksine benzer zevklere sahip kullanıcılar arasından atak kullanıcıların araştırılması yapılmıştır. Bu yüzden 5-9 ve 13-17 satırlarından da görüleceği üzere atak ve diğer kullanıcılar arasındaki korelasyon değerleri gerçek kullanıcılar arasında olduğunda daha küçük katsayılara sahip olmaktadır.

4.3. DBSCAN Algoritması ile Kümeleme

Atak kullanıcıların kendi aralarında ve normal kullanıcılar ile olan korelasyon katsayısı sıfıra yakındır. Korelasyon matrisi farklılık matrisine çevrildiği durumda atak kullanıcıların diğer noktalar ve kendi aralarındaki uzaklıklar normal kullanıcılara göre oldukça fazla olacaktır. Bir başka deyişle atak profiller gürültü noktaları gibi davranacaktır. Ardından Algoritma1 de sözde kodu verilen DBSCAN ile bu atak profiller tespit edilmiştir.

DBSCAN algoritması rastgele şekilleri kümelemek ve gürültü noktaları belirlemek için veri seti, komşuluk yarı çapı ve komşuluk yarı çapında olması gereken en az nokta olmak üzere üç girdiye ihtiyaç duyar. Bu girdilerin sonucunda veri setindeki noktaların küme etiketlerini ve gürültü noktalarını ortaya koyar. Algoritma 4.1'de DBSCAN algoritması verilmiştir. İkinci satırda gerekli veri kümesi, üçüncü satırda komşuluk yarı çapı, dördüncü satırda komşuluk içindeki gerekli en az nokta sayısı ve tüm noktaların sınıf etiketine 0 atanması ile algoritma başlar. Algoritma, veri kümesindeki her noktanın sınıf etiketi belirlenene kadar devam eder. Yedinci adımdaki if döngüsü ile ziyaret edilen

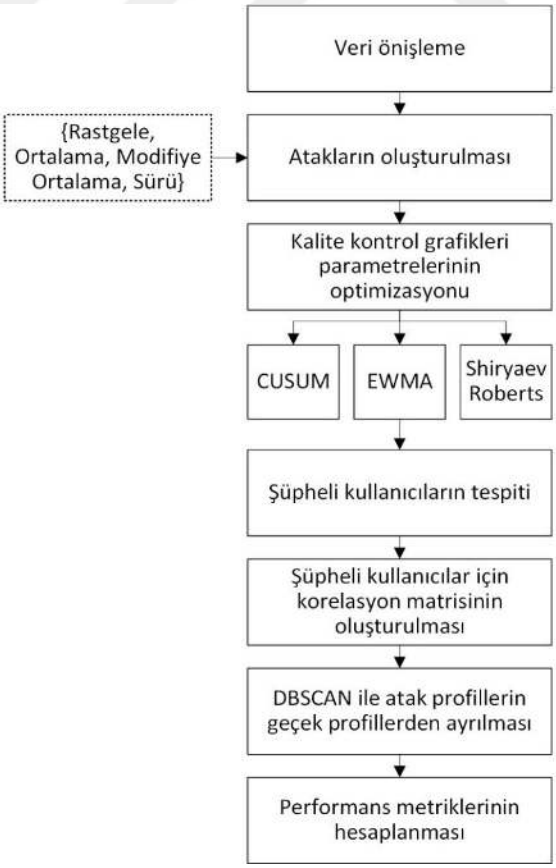
noktanın daha önceden ziyaret edilip edilmediği kontrol edilir. Eğer nokta daha önceden ziyaret edildi ise başka bir noktaya geçilir. Bu durumda ziyaret edilen noktaya 11. adımdaki if döngüsüne göre sınıf ataması yapılır. Eğer ε komşuluğunda ki nokta sayısı gerekli en az nokta sayısından daha küçükse nokta gürültü sınıfına atanır. Aksi durumda noktaya yeni bir sınıf etiketi atanır ve komşuluğundaki noktalar sorgulanarak bu sınıfa ait olup olmadığı belirlenir. Bu işlem tüm komşuluktaki noktalar sorgulanana kadar sınıf genişletilerek devam eder. Tüm noktaların ziyaret edilmesi ile algoritma sonlandırılır. Bu çalışmanın akış şeması Şekil 4.2'de verilmiştir.

Tablo 4.3. *Korelasyon ile kullanıcılar arasındaki uzaklıklar*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0,00	0,13	0,30	0,10	0,00	0,02	0,01	-0,01	-0,01	0,06	0,10	0,10	0,03	-0,01	-0,03	-0,03	0,04	0,17	0,14	0,18	0,05	0,30	0,10
2	0,13	0,00	0,22	0,11	-0,01	-0,01	-0,01	-0,03	-0,01	0,01	0,23	0,11	0,01	-0,02	-0,01	-0,01	0,08	0,36	0,28	0,07	0,09	0,21	0,25
3	0,30	0,22	0,00	0,14	0,01	-0,01	-0,02	0,02	-0,03	0,09	0,25	0,14	0,03	0,00	0,02	-0,03	0,02	0,26	0,18	0,31	0,03	0,32	0,20
4	0,10	0,11	0,14	0,00	-0,02	0,03	-0,01	0,02	0,03	0,07	0,15	0,15	0,01	0,02	-0,01	0,00	-0,02	0,07	0,07	0,15	0,02	0,28	0,17
5	0,00	-0,01	0,01	-0,02	0,00	-0,01	0,06	-0,01	-0,01	0,03	0,01	0,00	-0,01	-0,01	0,12	-0,01	-0,01	0,03	0,01	0,00	0,03	-0,01	0,02
6	0,02	-0,01	-0,01	0,03	-0,01	0,00	-0,01	-0,01	-0,01	-0,02	0,01	-0,05	-0,01	-0,01	-0,01	-0,01	-0,01	0,02	-0,01	0,02	-0,01	0,03	-0,01
7	0,01	-0,01	-0,02	-0,01	0,06	-0,01	0,00	-0,01	-0,01	0,03	-0,01	-0,01	-0,01	-0,01	0,06	-0,01	0,06	-0,03	-0,01	-0,02	0,03	-0,01	0,01
8	-0,01	-0,03	0,02	0,02	-0,01	-0,01	-0,01	0,00	-0,01	-0,02	0,03	0,00	-0,01	-0,01	0,06	0,06	-0,01	0,02	0,05	0,00	0,03	-0,02	0,02
9	-0,01	-0,01	-0,03	0,03	-0,01	-0,01	-0,01	-0,01	0,00	-0,02	-0,02	0,02	-0,01	-0,01	-0,01	-0,01	-0,01	-0,02	-0,01	0,01	-0,01	-0,02	-0,04
10	0,06	0,01	0,09	0,07	0,03	-0,02	0,03	-0,02	-0,02	0,00	0,13	0,22	0,01	-0,02	-0,02	-0,02	0,05	0,08	0,05	0,17	0,14	0,11	0,29
11	0,10	0,23	0,25	0,15	0,01	0,01	-0,01	0,03	-0,02	0,13	0,00	0,11	0,01	0,04	0,00	-0,01	0,00	0,21	0,16	0,23	0,06	0,27	0,33
12	0,10	0,11	0,14	0,15	0,00	-0,05	-0,01	0,00	0,02	0,22	0,11	0,00	-0,01	-0,02	0,00	0,00	0,03	0,10	0,05	0,09	0,13	0,15	0,18
13	0,03	0,01	0,03	0,01	-0,01	-0,01	-0,01	-0,01	-0,01	0,01	0,01	-0,01	0,00	-0,01	-0,01	-0,01	-0,01	-0,03	-0,01	0,03	0,04	0,03	0,03
14	-0,01	-0,02	0,00	0,02	-0,01	-0,01	-0,01	-0,01	-0,01	-0,02	0,04	-0,02	-0,01	0,00	-0,01	-0,01	-0,01	-0,02	-0,01	0,03	-0,01	-0,01	-0,02
15	-0,03	-0,01	0,02	-0,01	0,12	-0,01	0,06	0,06	-0,01	-0,02	0,00	0,00	-0,01	-0,01	0,00	-0,01	-0,01	-0,02	-0,01	0,01	-0,01	0,00	-0,01
16	-0,03	-0,01	-0,03	0,00	-0,01	-0,01	-0,01	0,06	-0,01	-0,02	-0,01	0,00	-0,01	-0,01	-0,01	0,00	-0,01	-0,01	-0,01	0,00	-0,01	0,01	-0,04
17	0,04	0,08	0,02	-0,02	-0,01	-0,01	0,06	-0,01	-0,01	0,05	0,00	0,03	-0,01	-0,01	-0,01	-0,01	0,00	0,11	0,00	-0,01	-0,01	0,02	0,03
18	0,17	0,36	0,26	0,07	0,03	0,02	-0,03	0,02	-0,02	0,08	0,21	0,10	-0,03	-0,02	-0,02	-0,01	0,11	0,00	0,34	0,12	0,14	0,18	0,28
19	0,14	0,28	0,18	0,07	0,01	-0,01	-0,01	0,05	-0,01	0,05	0,16	0,05	-0,01	-0,01	-0,01	-0,01	0,00	0,34	0,00	0,14	0,13	0,18	0,23
20	0,18	0,07	0,31	0,15	0,00	0,02	-0,02	0,00	0,01	0,17	0,23	0,09	0,03	0,03	0,01	0,00	-0,01	0,12	0,14	0,00	0,09	0,28	0,26
21	0,05	0,09	0,03	0,02	0,03	-0,01	0,03	0,03	-0,01	0,14	0,06	0,13	0,04	-0,01	-0,01	-0,01	-0,01	0,14	0,13	0,09	0,00	0,04	0,23
22	0,30	0,21	0,32	0,28	-0,01	0,03	-0,01	-0,02	-0,02	0,11	0,27	0,15	0,03	-0,01	0,00	0,01	0,02	0,18	0,18	0,28	0,04	0,00	0,26
23	0,10	0,25	0,20	0,17	0,02	-0,01	0,01	0,02	-0,04	0,29	0,33	0,18	0,03	-0,02	-0,01	-0,04	0,03	0,28	0,23	0,26	0,23	0,26	0,00

Algoritma 4.1. DBSCAN algoritması

```
function DBSCAN(D, ε, minPts)
Başlatma:
Veri Kümesi ←D
Uzaklık← ε
Minimum Nokta Sayısı← minPts
C←0
    for all P in D do
        if P için sınıf atandı then
            Sonraki P noktasına git
        else
            P'ye sınıf ata
            if ε komşuluğundaki nokta sayısı < minPts then
                P gürültü (aykırı) nokta
            else
                C←Yeni Sınıf
                Sınıfı Genişlet
            end if
        end if
    end for
end function
```



Şekil 4.2. *Önerilen online atak tespiti akış şeması*

4.4. Bilimsel Sonuçlar

Bu tez çalışmasında şüpheli kullanıcılar ÜAHO grafiği ile online olarak tespit edilmiş ardından şüpheli kullanıcılar arasından atak kullanıcıların tespiti için DBSCAN algoritması ile kümeleme işlemi gerçekleştirilmiştir. Çalışma da literatürde de sıklıkla kullanılan MovieLens100K ve MoviLens1M veri setleri çalışmanın doğruluğu için kullanılmıştır. İki veri setinin duyarlılık, kesinlik ve F1-skoru değerleri aşağıdaki tablolarda sunulmuştur.

Deneysel çalışmalarda %1, %3 ve %5 oranında atak büyüklükleri ve doldurma oranları kullanılmıştır. Rastgele, ortalama ve modifiye edilmiş ortalama atak olmak üzere üç farklı atak türü bahse konu veri kümelerine enjekte edilmiştir. Modifiye edilmiş ortalama atakta doldurma puanlamalar yapılırken ürünün puan ortalaması yerine puan dağılımı kullanılmıştır.

Bu çalışma kapsamında atak tespit süreci temel olarak üç bölümde incelenmiştir. Öncelikle farklı atakların üretilmesi ile süreç başlamaktadır. Ardında üretilen ataklar sisteme enjekte edilmiştir. İkinci bölümde üretilen ataklar ile gerçek kullanıcılar ile kalite kontrol grafikleri vasıtası ile tespit sürecine tabi tutulmaktadır. Ardından tespit edilen şüpheli kullanıcılar kümeleme yöntemi ile ayrıştırılarak atak kullanıcıların tespiti sağlanmaktadır. Tablo 4.4'te MovieLens100K veri kümesi için bahse konu bölümler için saniye cinsinden harcanan süreler verilmiştir. Tablodan görüleceği üzere rastgele ataklar için toplam süreç izafi olarak diğer ataklara göre daha az olmaktadır. Atağın bilgi düzeyi arttıkça atak için geliştirilen sürecin süreside artmaktadır.

Çalışmada kullanılan veri kümesinde filmlerin puan ortalamaları yüksektir. Bu yüzden itme türü ataklarda çekme türüne göre daha fazla şüpheli kullanıcı tespit edilmektedir. Bundan dolayı atak tespit süresi izafi olarak çekme türü ataklara göre uzun olmaktadır. Tablo 4.5'te MovieLens100K veri kümesine uygulanan farklı atak modelleri için süreler saniye cinsinden verilmiştir.

Tablo 4.6'da çekme türü ataklar için MovieLens1M veri kümesine uygulanan farklı atakların süreleri verilmiştir. Çekme türü ataklar için karşılaştırma yapıldığında daha büyük veri setlerinde uygulama sürelerinin uzadığı görülmektedir.

Tablo 4.4. *MovieLens100K veri kümesine uygulanan çekme türü atak için önerilen yöntemin süreleri (sn)*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Rastgele	Atak Üretme	0.0626	0.1732	0.2708	0.1610	0.4736	0.7738	0.2638	0.7692	1.2643
	Tespit	0.0004	0.0003	0.0003	0.0008	0.0008	0.0008	0.0017	0.0017	0.0017
	Kümeleme	0.0025	0.0025	0.0022	0.0024	0.0024	0.0026	0.0026	0.0027	0.0025
	Toplam	0.0655	0.1761	0.2733	0.1642	0.4769	0.7772	0.2681	0.7736	1.2685
Ortalama	Atak	0.0500	0.1463	0.2365	0.1384	0.4066	0.6950	0.2337	0.6971	1.1422
	Tespit	0.0003	0.0003	0.0003	0.0007	0.0007	0.0008	0.0016	0.0016	0.0016
	Kümeleme	0.0020	0.0024	0.0020	0.0023	0.0022	0.0023	0.0025	0.0025	0.0025
	Toplam	0.0523	0.1491	0.2388	0.1414	0.4096	0.6980	0.2378	0.7012	1.1463
M-Ortalama	Atak	0.0551	0.1671	0.2673	0.1550	0.4713	0.7776	0.2595	0.7708	1.2754
	Tespit	0.0004	0.0003	0.0003	0.0008	0.0008	0.0008	0.0017	0.0017	0.0017
	Kümeleme	0.0025	0.0023	0.0021	0.0023	0.0026	0.0023	0.0027	0.0026	0.0025
	Toplam	0.0579	0.1697	0.2697	0.1582	0.4747	0.7807	0.2639	0.7750	1.2796

Tablo 4.5. *MovieLens100K veri kümesine uygulanan itme türü atak için önerilen yöntemin süreleri (sn)*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Rastgele	Atak Üretme	0.0948	0.2666	0.4436	0.2571	0.7742	1.2941	0.4238	1.2811	2.1036
	Tespit	0.0032	0.0030	0.0030	0.0046	0.0046	0.0047	0.0067	0.0067	0.0066
	Kümeleme	0.0036	0.0185	0.0086	0.0129	0.0096	0.0122	0.0054	0.0062	0.0042
	Toplam	0.1015	0.2881	0.4552	0.2746	0.7884	1.3110	0.4359	1.2940	2.1143
Ortalama	Atak	0.1041	0.2649	0.4626	0.2745	0.7943	1.3099	0.4501	1.3237	2.2131
	Tespit	0.0030	0.0030	0.0030	0.0048	0.0047	0.0046	0.0068	0.0068	0.0068
	Kümeleme	0.0042	0.0040	0.0044	0.0056	0.0061	0.0037	0.0061	0.0058	0.0041
	Toplam	0.1114	0.2719	0.4699	0.2848	0.8051	1.3182	0.4630	1.3362	2.2241
M-Ortalama	Atak	0.1129	0.2726	0.4733	0.2796	0.8149	1.3530	0.4631	1.3585	2.2619
	Tespit	0.0035	0.0029	0.0029	0.0048	0.0047	0.0046	0.0069	0.0068	0.0068
	Kümeleme	0.0069	0.0043	0.0043	0.0044	0.0058	0.0045	0.0089	0.0069	0.0042
	Toplam	0.1233	0.2798	0.4806	0.2888	0.8254	1.3621	0.4788	1.3721	2.2729

Tablo 4.6. *MovieLens1M veri kümesine uygulanan çekme türü atak için önerilen yöntemin süreleri (sn)*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Rastgele	Atak Üretme	0.0135	0.0607	0.1508	0.0556	0.4240	1.1367	0.1277	1.1580	3.4175
	Tespit	0.0841	0.0777	0.0799	0.3247	0.2410	0.2547	0.5690	0.5322	0.4836
	Kümeleme	0.0207	0.0136	0.0145	0.0297	0.0288	0.0349	0.0497	0.0530	0.0405
	Toplam	0.1182	0.1520	0.2452	0.4100	0.6938	1.4263	0.7465	1.7433	3.9416
Ortalama	Atak	0.0134	0.0635	0.1453	0.0540	0.4270	1.1509	0.1257	1.1888	3.4750
	Tespit	0.0803	0.0904	0.0783	0.3157	0.2690	0.2527	0.5779	0.5266	0.4586
	Kümeleme	0.0221	0.0179	0.0141	0.0299	0.0294	0.0307	0.0423	0.0513	0.0411
	Toplam	0.1157	0.1718	0.2377	0.3996	0.7254	1.4344	0.7459	1.7667	3.9747
M-Ortalama	Atak	0.0354	0.1464	0.2869	0.1298	0.6564	1.6331	0.2525	1.6059	4.0618
	Tespit	0.0637	0.0947	0.0869	0.3390	0.2646	0.2766	0.5342	0.5082	0.4321
	Kümeleme	0.0250	0.0193	0.0150	0.0318	0.0294	0.0266	0.0451	0.0422	0.0360
	Toplam	0.1240	0.2604	0.3888	0.5005	0.9504	1.9364	0.8318	2.1563	4.5300

Benzer şekilde MovieLens1M veri kümesi için de itme türü ataklarda uygulama sürelerinin daha uzun olduğu Tablo 4.7'den görülebilir.

Tablo 4.7. *MovieLens1M veri kümesine uygulanan itme türü atak için önerilen yöntemin süreleri (sn)*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Rastgele	Atak Üretme	0.0118	0.0489	0.1023	0.0355	0.2504	0.6387	0.0806	0.6289	1.6884
	Tespit	0.2511	0.2843	0.2601	0.4476	0.3762	0.3048	0.6119	0.5533	0.5301
	Kümeleme	0.0181	0.0212	0.0164	0.0253	0.0227	0.0209	0.0315	0.0301	0.0283
	Toplam	0.2811	0.3544	0.3788	0.5084	0.6493	0.9644	0.7240	1.2122	2.2468
Ortalama	Atak	0.0111	0.0479	0.1023	0.0352	0.2466	0.6329	0.0806	0.6344	1.7097
	Tespit	0.2372	0.2808	0.2664	0.4407	0.3882	0.3011	0.5890	0.5649	0.5150
	Kümeleme	0.0179	0.0201	0.0167	0.0274	0.0225	0.0213	0.0306	0.0297	0.0295
	Toplam	0.2662	0.3488	0.3854	0.5033	0.6573	0.9553	0.7002	1.2290	2.2542
M-Ortalama	Atak	0.0217	0.0837	0.1539	0.0668	0.3218	0.7913	0.1273	0.7503	1.8215
	Tespit	0.2508	0.2876	0.2615	0.4595	0.3511	0.3147	0.6225	0.5058	0.4674
	Kümeleme	0.0181	0.0221	0.0174	0.0249	0.0224	0.0208	0.0318	0.0285	0.0277
	Toplam	0.2907	0.3934	0.4328	0.5512	0.6953	1.1268	0.7817	1.2846	2.3167

Tablo 4.8'de MoviLens100K veri setine enjekte edilmiş çekme atak türünün metriklerden aldığı değerler verilmiştir. Doldurma oranının %1 olduğu durumların haricinde %99 un üzerinde başarılar elde edilmiştir. Doldurma oranın düşük olduğu testlerde ise bu düşüklük DBSCAN algoritmasının parametre ayarlaması ile çözülebilmektedir.

Tablo 4.8. *MovieLens100K veri setine enjekte edilmiş çekme atak türü performans ölçütleri sonuçları*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Duyarlılık	Rastgele	0.9960	1.0000	1.0000	0.9925	0.9998	1.0000	0.9859	0.9996	1.0000
	Ortalama	0.9762	0.9996	1.0000	0.9947	0.9980	0.9997	0.9847	0.9971	0.9996
	M-Ortalama	0.9802	0.9992	1.0000	0.9961	0.9980	0.9999	0.9846	0.9978	0.9999
Kesinlik	Rastgele	0.9837	0.9810	0.9849	0.9883	0.9898	0.9892	0.9923	0.9910	0.9921
	Ortalama	0.9836	0.9810	0.9849	0.9838	0.9898	0.9893	0.9880	0.9910	0.9921
	M-Ortalama	0.9832	0.9817	0.9855	0.9836	0.9910	0.9906	0.9872	0.9918	0.9918
F1-Skoru	Rastgele	0.9898	0.9904	0.9924	0.9904	0.9948	0.9946	0.9891	0.9953	0.9960
	Ortalama	0.9799	0.9902	0.9924	0.9892	0.9939	0.9945	0.9863	0.9941	0.9958
	M-Ortalama	0.9817	0.9904	0.9927	0.9898	0.9945	0.9952	0.9859	0.9948	0.9958

Benzer bir şekilde, atakların tespiti için MovieLens1M veri setinde de yüksek duyarlılık, kesinlik ve F1-Skor değerleri elde edilmiştir. MovieLens100K veri setinden farklı olarak düşük doldurma oranlarında %99 üzerinde başarılar elde edilmiştir. Sonuç olarak %99,5 in üzerinde F1 değerleri bu veri seti için elde edilmiştir. Sonuçların tamamı Tablo 4.9'da ayrıntılı olarak verilmiştir. MovieLens1M veri setinde MovieLens100K veri setine göre daha iyi sonuçlar elde edilmiştir. Bu durum göstermiştir ki, eğer veri setinin boyutu büyürse daha iyi sonuçlar elde edilecektir.

Çekme türü atak modellerinde olduğu gibi itme türü atak modelleri de aynı doldurma ve atak boyutlarında veri setlerine enjekte edilmiştir. Bu çalışmada atakların enjekte edildiği ürünlerin ortalamaları yüksek olduğu için itme türü atak modellerini tespit etmek daha zor olmuştur. Bu yüzden çekme türü atak modellerine göre göreceli olarak başarı oranlarında düşüşler gözlenmiştir. Buna rağmen %99'un üzerindeki başarılar Tablo 4.10'da görüldüğü gibi elde edilmiştir.

Tablo 4.9. *MovieLens1M veri setine enjekte edilmiş çekme atak türü performans ölçütleri sonuçları*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Duyarlılık	Rastgele	0.9998	0.9999	0.9999	0.9999	1.0000	1.0000	0.9999	1.0000	1.0000
	Ortalama	0.9997	0.9999	0.9998	0.9999	1.0000	1.0000	0.9999	1.0000	1.0000
	M-Ortalama	0.9996	0.9999	0.9999	0.9999	0.9999	1.0000	0.9999	1.0000	1.0000
Kesinlik	Rastgele	0.9898	0.9916	0.9915	0.9975	0.9979	0.9978	0.9983	0.9986	0.9985
	Ortalama	0.9895	0.9915	0.9914	0.9974	0.9977	0.9977	0.9984	0.9987	0.9987
	M-Ortalama	0.9895	0.9924	0.9919	0.9975	0.9977	0.9979	0.9984	0.9986	0.9987
F1-Skoru	Rastgele	0.9947	0.9957	0.9957	0.9987	0.9989	0.9989	0.9991	0.9993	0.9993
	Ortalama	0.9946	0.9957	0.9956	0.9986	0.9989	0.9988	0.9991	0.9993	0.9993
	M-Ortalama	0.9945	0.9961	0.9959	0.9987	0.9988	0.9989	0.9991	0.9993	0.9993

İzafi olarak MovieLens1M veri setinde daha başarılı sonuçlar elde edilmiştir. Tablo 4.11'den de görüleceği üzere bazı testlerde %100 oranında duyarlılık değerleri elde edilmiştir. Bu duyarlılık değerleri elde edilirken kesinlik değerlerinden ödün verilmemiş bu durumda F1 değerlerine yansımıştır.

Tablo 4.10. *MovieLens100K veri setine enjekte edilmiş itme atak türü performans ölçütleri sonuçları*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Duyarlılık	Rastgele	0.9990	0.9998	1.0000	0.9903	0.9998	1.0000	0.9839	0.9999	0.9999
	Ortalama	0.9712	0.9956	0.9980	0.9939	0.9953	0.9988	0.9798	0.9941	0.9986
	M-Ortalama	0.9862	0.9828	0.9982	0.9850	0.9922	0.9957	0.9860	0.9922	0.9953
Kesinlik	Rastgele	0.9756	0.9754	0.9761	0.9910	0.9908	0.9912	0.9941	0.9943	0.9941
	Ortalama	0.9717	0.9840	0.9846	0.9740	0.9940	0.9937	0.9832	0.9966	0.9966
	M-Ortalama	0.9721	0.9926	0.9922	0.9745	0.9972	0.9972	0.9798	0.9984	0.9984
F1-Skoru	Rastgele	0.9871	0.9875	0.9879	0.9906	0.9953	0.9956	0.9889	0.9971	0.9970
	Ortalama	0.9714	0.9898	0.9912	0.9838	0.9947	0.9963	0.9815	0.9954	0.9976
	M-Ortalama	0.9791	0.9877	0.9952	0.9797	0.9947	0.9965	0.9829	0.9953	0.9968

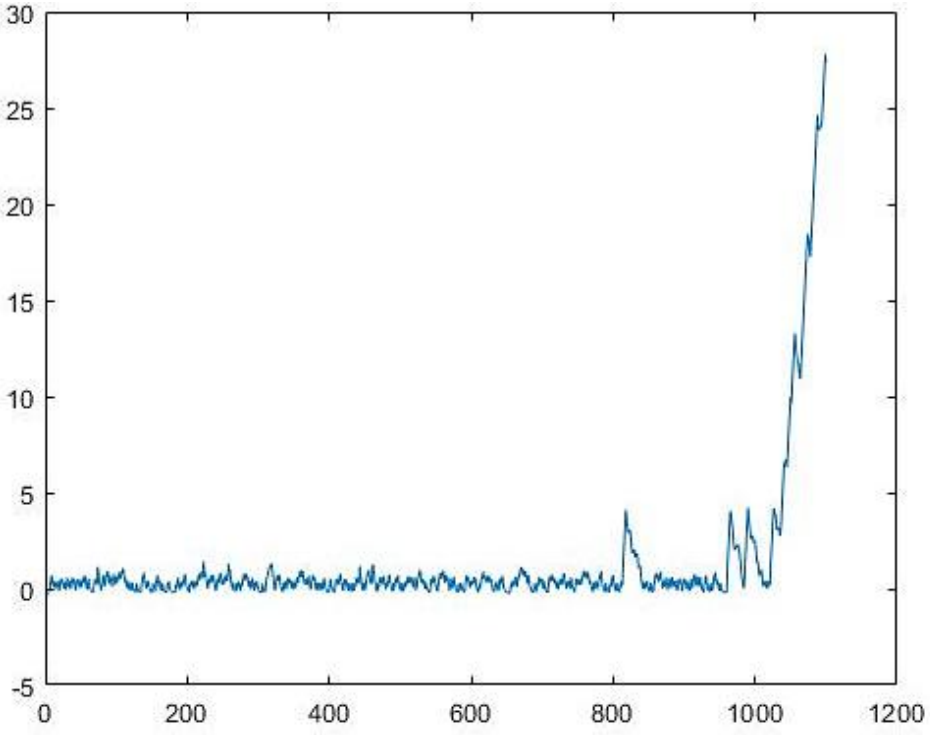
Tablo 4.11. *MovieLens1M veri setine enjekte edilmiş itme atak türü performans ölçütleri sonuçları*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Duyarlılık	Rastgele	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Ortalama	0.9999	1.0000	1.0000	0.9998	1.0000	1.0000	0.9998	1.0000	1.0000
	M-Ortalama	0.9999	1.0000	1.0000	0.9999	1.0000	1.0000	0.9998	1.0000	1.0000
Kesinlik	Rastgele	0.9888	0.9889	0.9891	0.9964	0.9964	0.9964	0.9978	0.9978	0.9978
	Ortalama	0.9889	0.9888	0.9891	0.9964	0.9964	0.9964	0.9979	0.9978	0.9978
	M-Ortalama	0.9890	0.9890	0.9890	0.9963	0.9963	0.9964	0.9978	0.9978	0.9978
F1-Skoru	Rastgele	0.9944	0.9944	0.9945	0.9982	0.9982	0.9982	0.9989	0.9989	0.9989
	Ortalama	0.9944	0.9944	0.9945	0.9981	0.9982	0.9982	0.9988	0.9989	0.9989
	M-Ortalama	0.9944	0.9945	0.9945	0.9981	0.9982	0.9982	0.9988	0.9989	0.9989

Bu çalışmada, üç farklı atak türü üç farklı doldurma oranı ve üç farklı ata boyutunda literatürde sıklıkla kullanılan iki veri setine enjekte edilmiştir. Test çalışmaların sonuçları göstermiştir ki, online atak tespit yöntemleri başarılı bir şekilde öneri sistemlerine uygulanabilir. Kalite ölçütlerinin ufak bir kısmında %99'un altında başarılar elde edilmiş olsa da test çalışmaların büyük bir kısmında %99 un üzerinde başarılar elde edilmiştir.

4.5. Shiryaev-Roberts Prosedürü ile Süreç Kontrolü

SR prosedürünün geçerliliği için MovieLens1M veri seti üzerinde bir deneysel çalışma gerçekleştirilmiştir. Bahsedilen veri setinde rasgele seçilmiş bir ürüne karma bir atak modeli enjekte edilmiştir. Enjekte edilen atak modeli içinde rasgele, ortalama ve modifiye edilmiş ortalama atak modelleri bulunmaktadır. Varsayılan ortalamadaki değişim bir standart sapma $\delta = 1$ boyutunda ve standart satma ise 1.5 $\sigma_1 = 1.5 * \sigma_0$kat olarak alınmıştır. Şekil 4.3'de de görüldüğü üzere 813-817, 960-964, 985-989, ve 1021-1099 arasında atak profilleri eklendiğinde SR istatistiğinde artışlar gözlemlenmiştir. Tespit edilen şüpheli kullanıcılar ayrıntılı incelendiğinde 1021 ile 1099 arasındaki profillerde gerçek kullanıcılarında bulunduğu saptanmış ve bahse konu kullanıcılar ÜAHO yönteminde olduğu gibi DBSCAN yöntemi ile ayrıştırılmıştır.



Şekil 4.3. *Atak Enjekte Edilen Durumda SR İstatistiği*

4.6. Çokyüzlü Konik Fonksiyonlar ve Aşırı Örnekleme

Bu çalışmada, modeli doğrulamak için MovieLens100K veri seti kullanılmıştır. Bu veri seti, 943 kullanıcıdan 1682 öğeye kadar 100.000 derecelendirmeden oluşur. Bu 943 kullanıcı gerçek kullanıcılardır ve bu veri kümesine 50 atak kullanıcı eklenmiştir. Bu çalışma, sentetik kullanıcıların ürünlerin derecelendirme dağılımına zarar vermesinden dolayı sisteme bir sonraki giren kullanıcıların gerçek olup olmadığını belirlemeyi amaçlamaktadır.

Mevcut tablo, veri setindeki 943 gerçek kullanıcıyı ve 50 sentetik kullanıcıyı göstermektedir. Veri setinin temel formunda konik ayırma uygulandığında yaklaşık %94 doğruluk oranı elde edilmektedir. Sınıf üyeleri göz önüne alındığında, bu veri kümesi için bir sınıf dengesizliği sorunu vardır.

Bahsedilen sorunun üstesinden gelmek için artan sentetik kullanıcı sayısı oranları ile ODBOT uygulanmaktadır. Test ve öğrenme setlerinin doğruluk oranları Tablo 4.12'de verilmiştir. Tablo 4.12'den de görülebileceği gibi, öğrenme doğruluk oranları her zaman 1'e eşittir. Öte yandan, aşırı örnekleme seviyeleri ile test doğruluk oranları artmaktadır. Bu çalışmada, önerilen yöntemin etkinliğini göstermek için on kat çapraz doğrulama uygulanmıştır.

Tablo 4.12. *Farklı Aşırı Örnekleme Seviyeleri için Doğruluk Oranları*

Aşırı Örnekleme Seviyesi	Öğrenme Kümesi için Doğruluk Oranları	Test Kümesi için Doğruluk Oranları
%0	1	0.9396
%10	1	0.9430
%20	1	0.9476
%30	1	0.9476
%40	1	0.9547
%50	1	0.9564
%60	1	0.9609
%70	1	0.9620
%80	1	0.9662
%90	1	0.9675

5. AŞIRI ÖRNEKLEME TABANLI DESTEK VEKTÖR MAKİNELERİ İLE ŞİLİN ATAK TESPİTİ

Bu çalışmada, hesaplamalı değerlendirme için MovieLens100K veri kümesi kullanılmıştır. Bu veri kümesinde 943 kullanıcıdan 1652 film hakkında 100.000 değerlendirme bulunmaktadır. Her kullanıcı en az 20 filme puan vermektedir. Bu veri kümesi yüksek seyrekliğe sahiptir ve seyreklik oranı yaklaşık %93,7'dir. MovieLens100K veri kümesi 5 yıldızlı derecelendirmeye sahiptir. Puanlar {1,2,3,4,5} kümesinde değerler alır. Veri kümesi dört sütunlu veriler içerir. Kullanıcı ve film kimlikleri sırasıyla birinci ve ikinci sütunlarda verilmiştir. Verilen oranlar üçüncü sütunda yer almaktadır. Verilen oranların zaman damgaları dördüncü sütuna yerleştirilmiştir.

Bu çalışmada şilin ataklarının tespiti için iki DVM modeli kullanılmıştır. Yöntemlerin karşılaştırılmasında kullanılmak için aşırı örnekleme kullanılan model ODBOT-DVM, kullanılmayan model ise temel DVM (T-DVM) olarak adlandırılacaktır.

Bu çalışmada 4 farklı atak modeli ve her atak modeli için itme ve çekme atak türleri veri kümelerine enjekte edilmiştir. Enjekte edilen atakların tespiti için her örneklemde ODBOT-DVM ve T-DVM ile modelleri üretilmiştir. Üretilen modellerin geçerliliğini göstermek için 10 kez çapraz doğrulama kullanılmıştır. Bahsedilen ataklar kullanılan MovieLens100K veri kümesindeki en yüksek puanlama sayısına sahip ilk 50 ürüne enjekte edilmiştir. 10 kez çapraz doğrulamada 943 kullanıcı 10 parçaya bölünmüştür. Her parça ayrı ayrı test kümesi olarak kullanılırken geri kalan dokuz parça öğrenme kümesi olarak kullanılmıştır. Her ürün için incelenen farklı atak ve dolgu büyüklükler için 10'ar adet atak üretilmiştir. Bu atak kümelerinden bir tanesi öğrenme kümesinde kullanılırken kendisi de dahil olmak üzere geri kalan dokuz atak kümesi test için kullanılmıştır. Bu doğrultuda her ürün için her atak ve dolgu büyüklüğünde 100 tekrar sağlanmıştır. Sonuç olarak her farklı atak ve dolgu büyüklüğü için 5000 ODBOT-DVM ve T-DVM modelleri kurulmuştur.

Literatürdeki en basit atak türü olan rastgele ataklar, dolgu öğeleri için genel ortalamayı kullanmaktadır. Tablo 5.1'da itme türü rastgele ataklar için farklı atak ve dolgu büyüklükleri için aşırı örnekleme kullanıldığı ve kullanılmadığı durumlarda kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları sunulmuştur. Tabloda aşırı örnekleme kullanılan ve kullanılmayan DVM modellerinin sonuçları yan yana verilerek daha iyi sonuçlar kalın yazı tipi ile belirtilmiştir. Kesinlik sonuçlarına bakıldığında T-

DVM modelinin daha iyi sonuçlar ürettiği gözlemlenmiştir. Fakat aşırı örneklemenin kullanıldığı sonuçlar ile arasında anlamlı bir fark bulunmamaktadır. Kesinlik sonuçları değerlendirildiğinde iki model arasında binde üçlük fark bulunmaktadır. Öte yandan duyarlılık sonuçları değerlendirildiğinde ODBOT-DVM modeli ile T-DVM model arasında yaklaşık olarak yüzde yirmi beşlik bir fark bulunmaktadır. Kesinlik ve duyarlılık sonuçlarının bir fonksiyonu olan F1-skoruna bakıldığında yine ODBOT-DVM modelinin daha iyi sonuçlar ürettiğini görmekteyiz. Sadece iki örneklemde T-DVM yöntemi daha iyi sonuçlar vermiştir. Fakat ODBOT-DVM yönteminin sonuçları arasında anlamlı bir fark gözlemlenmemiştir. Doğruluk sonuçları değerlendirildiğinde F1-skoru ile paralel sonuçlar elde edilmiştir. Doğruluk sonuçları arasındaki fark, F1-skoru sonuçları arasındaki farktan daha azdır. Bu farklılığın sebebi sınıflar arasında ciddi oranda fark olmasından kaynaklanmaktadır. Yöntemler gerçek kullanıcıları yüksek doruluk oranı ile sınıflandırmaktadır. Veri setinde gerçek kullanıcılar ile atak kullanıcılar arasında yüz kata varan fark bulunduğu için doğruluk oranlarındaki fark daha az olmaktadır. Bir başka deyişle doğruluk ve F1-skoru metrikleri yanlış pozitif ve yanlış negatif sonuçları için farklı önem derecelerine sahiptir. Gerçek ve atak kullanıcı sayılarının eşit dağılım gösterdiği veri kümelerinde, bahsedilen iki metriğin birbirine yakın sonuçlar üretecektir. MovieLens100K verisi için enjekte edilen itme türü rastgele ataklar için sonuçlar değerlendirildiğinde ODBOT-DVM yöntemi ile elde edilen model anlamlı olarak daha iyi sonuçlar üretmiştir. Özellikle duyarlılık ve F1-skorları için sonuçlar arasında büyük farklar elde edilmiştir.

Tablo 5.1. *İtme türü rasgele ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skor		Doğruluk	
AO	DO	RA_İtme ODBOT	RA_İtme	RA_İtme ODBOT	RA_İtme	RA_İtme ODBOT	RA_İtme	RA_İtme ODBOT	RA_İtme
1%	1%	0,9935	**0,9998**	**0,9983**	0,8029	**0,9959**	0,8906	**0,9992**	0,9811
1%	3%	0,9925	**0,9996**	**0,9790**	0,1038	**0,9857**	0,1880	**0,9973**	0,9141
1%	5%	0,9920	**0,9990**	**0,9375**	0,1007	**0,9640**	0,1829	**0,9933**	0,9138
3%	1%	0,9977	**0,9990**	**0,9999**	0,9993	0,9988	**0,9992**	0,9994	**0,9996**
3%	3%	0,9966	**0,9993**	**0,9917**	0,8711	**0,9941**	0,9308	**0,9972**	0,9687
3%	5%	**0,9950**	0,9939	**0,9907**	0,7577	**0,9928**	0,8599	**0,9965**	0,9404
5%	1%	0,9983	**0,9990**	**1,0000**	1,0000	0,9991	**0,9995**	0,9994	**0,9996**
5%	3%	0,9973	**0,9990**	**0,9984**	0,9858	**0,9978**	0,9923	**0,9985**	0,9947
5%	5%	0,9960	**0,9950**	**0,9982**	0,9677	**0,9971**	0,9812	**0,9980**	0,9871

Veri kümesine enjekte edilen çekme türü rastgele ataklar için performans metrikleri sonuçları Tablo 5.2'da sunulmuştur. Kesinlik sonuçları değerlendirildiğinde ODBOT-DVM modeli bir örneklemde daha iyi sonuç üretmiştir. Kesinlik sonuçlarının ortalaması değerlendirildiğinde iki model sonuçları arasında yaklaşık olarak %1,5 fark bulunmaktadır. Duyarlılık sonuçları değerlendirildiğinde tüm örneklemlerde ODBOT-DVM modeli daha iyi sonuçlar üretmiştir. ODBOT-DVM modeli T-DVM modeline göre ortalamada yaklaşık olarak %40 daha iyi sonuçlar üretmiştir. F1-skoru için ODBOT-DVM modelinin sonuçları ortalamada %33 daha iyidir. Doğruluk için de F1-skoru ile aynı yorumları yapmak mümkündür. Fakat burada iyileşme oranı %6 ya düşmektedir. MovieLens100K verisi için enjekte edilen çekme türü rastgele ataklar için sonuçlar değerlendirildiğinde ODBOT-DVM yöntemi ile elde edilen model anlamlı olarak daha iyi sonuçlar üretmiştir.

Tablo 5.2. *Çekme türü rasgele ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skor		Doğruluk	
AO	DO	RA_Çekme ODBOT	RA_Çekme	RA_Çekme ODBOT	RA_Çekme	RA_Çekme ODBOT	RA_Çekme	RA_Çekme ODBOT	RA_Çekme
1%	1%	0,9699	**0,9998**	**0,9828**	0,1302	**0,9763**	0,2305	**0,9954**	0,9166
1%	3%	0,9647	**0,9994**	**0,8116**	0,1001	**0,8816**	0,1820	**0,9791**	0,9137
1%	5%	0,9546	**0,9970**	**0,6660**	0,1000	**0,7846**	0,1818	**0,9649**	0,9137
3%	1%	0,9894	**0,9952**	**0,9930**	0,9566	**0,9912**	0,9755	**0,9957**	0,9884
3%	3%	0,9876	**0,9938**	**0,9159**	0,4069	**0,9504**	0,5774	**0,9769**	0,8562
3%	5%	**0,9857**	0,9814	**0,9424**	0,4285	**0,9636**	0,5966	**0,9828**	0,8601
5%	1%	0,9929	**0,9951**	**0,9986**	0,9956	**0,9957**	0,9954	**0,9970**	0,9968
5%	3%	0,9916	**0,9959**	**0,9697**	0,8148	**0,9805**	0,8963	**0,9867**	0,9347
5%	5%	0,9906	**0,9930**	**0,9812**	0,8442	**0,9859**	0,9126	**0,9903**	0,9439

Sınıflar arası örneklem sayılarının ciddi farklılıklar gözlemlenen veri kümeleri için önerdiğimiz ODBOT-DVM yöntemi rastgele ataklar için anlamlı derecede farklılıklar üreten sonuçlar vermiştir. İtme türü rastgele ataklarda daha yüksek sonuçlar elde edilirken çekme türü rastgele ataklarda iyileşme oranı daha yüksek olmuştur.

Rastgele ataklardan farklı olarak ortalama ataklarda dolgu ürünlerinin puanlamaları dolgu ürünlerinin puan ortalamaları ile yapılmaktadır. Rastgele atak modelline göre her bir ürünün ayrı ayrı puan ortalaması gerektirmektedir. İtme türü ortalama ataklar için performans metrikleri sonuçları Tablo 5.3'de sunulmuştur. Kesinlik metriği için T-DVM

modeli ile binde üç daha iyi sonuçlar elde edilmiştir. Duyarlılık metriğinde iki deney kümesinde %100 başarı elde edilmiştir. Diğer sonuçlar değerlendirildiğinde ortalama olarak ODBOT-DVM modeli ile %22 daha iyi sonuçlar elde edilmiştir. F1-skorunda T-DVM modeli ile iki deney kümesinde izafi olarak daha iyi sonuç elde edilmiştir. Fakat anlamlı bir fark gözlemlenmemiştir. Kalan yedi deney kümesinde ODBOT-DVM modeli ile daha iyi sonuçlar üretilmiştir. Son olarak doğruluk metriği için F1-skoru için yapılan yorumlar geçerlidir. Yine rastgele atak modelinde olduğu gibi F1-skoruna göre iyileştirme oranı daha düşük kalmıştır. MovieLens100K verisi için enjekte edilen itme türü ortalama ataklar için sonuçlar değerlendirildiğinde ODBOT-DVM yöntemi ile elde edilen model anlamlı olarak daha iyi sonuçlar üretmiştir.

Tablo 5.3. *İtme türü ortalama ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skor		Doğruluk	
AO	DO	AAPush ODBOT	OA_İtme	OA_İtme ODBOT	OA_İtme	OA_İtme ODBOT	OA_İtme	OA_İtme ODBOT	OA_İtme
1%	1%	0,9931	**0,9991**	**0,9999**	0,9813	**0,9965**	0,9901	**0,9993**	0,9981
1%	3%	0,9918	**0,9999**	**0,9838**	0,1381	**0,9878**	0,2427	**0,9977**	0,9174
1%	5%	0,9905	**0,9933**	**0,9522**	0,1039	**0,9710**	0,1882	**0,9945**	0,9140
3%	1%	0,9972	**0,9983**	**1,0000**	**1,0000**	0,9986	**0,9992**	0,9993	**0,9996**
3%	3%	0,9959	**0,9996**	**0,9960**	0,9454	**0,9959**	0,9717	**0,9980**	0,9867
3%	5%	0,9938	**0,9944**	**0,9943**	0,8289	**0,9940**	0,9041	**0,9971**	0,9576
5%	1%	0,9979	**0,9986**	**1,0000**	**1,0000**	0,9990	**0,9993**	0,9993	**0,9995**
5%	3%	0,9964	**0,9992**	**0,9992**	0,9950	**0,9978**	0,9971	**0,9985**	0,9980
5%	5%	0,9955	**0,9961**	**0,9989**	0,9787	**0,9972**	0,9873	**0,9981**	0,9913

Çekme türü ortalama ataklar için performans metrikleri sonuçları Tablo 5.4'te verilmiştir. Kesinlik metriği için DVM-T tüm deney kümelerinde yaklaşık olarak %1 daha iyi sonuçlar üretmiştir. Duyarlılık metriği sonuçlarına bakıldığında ODBOT-DVM %44 oranında daha iyi sonuçlar elde ettiği gözlemlenmektedir. F1-skoru için bir deney kümesinde aynı değer elde edilirken, kalan sekiz deney kümesinde ODBOT-DVM ile daha iyi sonuçlar elde edilmiştir. Benzer sonuçlar doğruluk metriği için de elde edilmiştir. MovieLens100K verisi için enjekte edilen çekme türü ortalama ataklar için sonuçlar değerlendirildiğinde ODBOT-DVM yöntemi ile elde edilen model anlamlı olarak daha iyi sonuçlar üretmiştir.

Tablo 5.4. *Çekme türü ortalama ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skor		Doğruluk	
AO	DO	OA_Çekme ODBOT	OA_Çekme	OA_Çekme ODBOT	OA_Çekme	OA_Çekme ODBOT	OA_Çekme	OA_Çekme ODBOT	OA_Çekme
1%	1%	0,9694	**0,9968**	**0,9899**	0,3969	**0,9795**	0,5677	**0,9960**	0,9421
1%	3%	0,9642	**0,9990**	**0,8509**	0,1015	**0,9040**	0,1843	**0,9827**	0,9138
1%	5%	0,9577	**0,9799**	**0,7077**	0,1004	**0,8139**	0,1821	**0,9690**	0,9136
3%	1%	0,9891	**0,9922**	**0,9985**	0,9934	**0,9938**	0,9928	**0,9970**	0,9965
3%	3%	0,9878	**0,9981**	**0,9190**	0,5380	**0,9522**	0,6992	**0,9777**	0,8883
3%	5%	0,9856	**0,9841**	**0,9360**	0,4567	**0,9601**	0,6239	**0,9812**	0,8671
5%	1%	0,9928	**0,9932**	**0,9997**	0,9992	**0,9962**	**0,9962**	**0,9974**	**0,9974**
5%	3%	0,9920	**0,9978**	**0,9697**	0,8677	**0,9807**	0,9282	**0,9868**	0,9535
5%	5%	0,9907	**0,9931**	**0,9768**	0,8281	**0,9837**	0,9031	**0,9888**	0,9384

ODBOT-DVM ile elde edilen sonuçlara göre çekme türü ortalama ataklarda kesinlik, F1-skor ve doğruluk metriklerinde daha iyi sonuçlar elde edilirken, duyarlılık metriği için itme türü ortalama ataklarda daha iyi sonuçlar elde edilmiştir.

Bu çalışma kapsamında literatürde var olan ortalama ataklar modifiye edilerek daha etkin hale getirilmiş ardından veri kümesine enjekte edilmiştir. Ortalama ataklar dolgu ürünler için ürünün kendi ortalamasını kullanırken modifiye edilmiş ortalama ataklar her ürünün puanlama dağılımını elde ederek, dolgu ürünleri için kendi dağılımından rastgele elde edilmiş puanlar ile puanlanmaktadır. Bu sayede dolgu ürünlerinin puanlamalarında çeşitlilik elde edilerek atağın tespiti güçleştirilmiştir. İtme türü modifiye edilmiş ortalama ataklar için performans metrik sonuçları Tablo 5.5'te verilmiştir. Kesinlik metriği için T-DVM modeli ile yaklaşık olarak %1,4 daha iyi sonuçlar elde edilmiştir. Duyarlılık metriği için tüm deney kümelerinde ODBOT-DVM yöntemi ile yaklaşık olarak %35 daha iyi sonuçlar elde edilmiştir. Fakat atak büyüklüğü %1, dolgu boyutu %5 deney kümesine bakıldığında, duyarlılık sonucunun yaklaşık olarak %78 olarak elde edildiği gözlemlenmiştir. Diğer atak modelleri içindeki en düşük oranlardan birisi bu atak modeli için gerçekleşmiştir. Buda atağın tespitinin zorlaştığını göstermektedir. F1-skoru için iki deney kümesi haricinde ODBOT-DVM ile daha iyi sonuçlar elde edilmiştir. F1-skorunda ortalama olarak ODBOT-DVM modeli ile yaklaşık olarak %30 daha iyi sonuçlar elde edilmiştir. Doğruluk metriği için F-skoru için yapılan yorumlar geçerlidir. Fakat doğrulama metriğindeki iyileştirme oranı %5 ile sınırlı kalmıştır. MovieLens100K verisi için enjekte edilen itme türü modifiye edilmiş ortalama ataklar için sonuçlar

değerlendirildiğinde ODBOT-DVM yöntemi ile elde edilen model anlamlı olarak daha iyi sonuçlar üretmiştir.

Tablo 5.5. *İtme türü modifiye edilmiş ortalama ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skor		Doğruluk	
AO	DO	MOA_İtme ODBOT	OA_İtme	MOA_İtme ODBOT	OA_İtme	MOA_İtme ODBOT	MOA_İtme	MOA_İtme ODBOT	MOA_İtme
1%	1%	0,9693	**0,9982**	**0,9949**	0,3870	**0,9819**	0,5577	**0,9965**	0,9412
1%	3%	0,9660	**0,9996**	**0,8986**	0,1002	**0,9311**	0,1821	**0,9872**	0,9137
1%	5%	0,9608	**0,9974**	**0,7794**	0,1000	**0,8607**	0,1818	**0,9758**	0,9137
3%	1%	0,9892	**0,9932**	**0,9993**	0,9964	0,9942	**0,9948**	0,9972	**0,9975**
3%	3%	0,9886	**0,9989**	**0,9338**	0,5590	**0,9604**	0,7168	**0,9814**	0,8934
3%	5%	**0,9867**	0,9863	**0,9365**	0,4612	**0,9609**	0,6285	**0,9816**	0,8684
5%	1%	0,9929	**0,9942**	**0,9995**	0,9989	0,9962	**0,9966**	0,9974	**0,9976**
5%	3%	0,9928	**0,9985**	**0,9751**	0,8912	**0,9839**	0,9418	**0,9889**	0,9618
5%	5%	0,9915	**0,9934**	**0,9748**	0,8516	**0,9831**	0,9171	**0,9884**	0,9466

Çekme türü modifiye edilmiş ortalama atak türleri için performans metrikleri sonuçları Tablo 5.6'da verilmiştir. Kesinlik metriği için bakıldığında ortalama olarak %1,3 T-DVM ile dağa doğru sonuçlar elde edilmiştir. Duyarlılık metriğinde ise ODBOT-DVM ile yaklaşık %35 daha iyi sonuçlar elde edilmiştir. Yine burada %78 gibi düşük bir oran elde edilmiş ama temel modelde elde edilen sonuç %10 da kalmıştır. F1-skor değerlerine göre iki deney kümesi için T-DVM ile daha iyi sonuçlar elde edilirken geri kalan yedi deney kümesi için ODBOT-DVM ile daha iyi sonuçlar elde edilmiştir. Ortalama olarak ODBOT-DVM modeli ile %28 daha iyi sonuçlar elde edilmiştir. Son olarak doğruluk metriği için yine ODBOT-DVM ile daha iyi sonuçlar elde edilirken buradaki iyileşme oranı %5 ile sınırlı kalmıştır. MovieLens100K verisi için enjekte edilen çekme türü modifiye edilmiş ortalama ataklar için sonuçlar değerlendirildiğinde ODBOT-DVM yöntemi ile elde edilen model anlamlı olarak daha iyi sonuçlar üretmiştir.

Tablo 5.6. *Çekme türü modifiye ortalama ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skoru		Doğruluk	
AO	DO	MOA_Çekme ODBOT	MOA_Çekme	MOA_Çekme ODBOT	MOA_Çekme	MOA_Çekme ODBOT	MOA_Çekme	MOA_Çekme ODBOT	MOA_Çekme
1%	1%	0,9693	**0,9982**	**0,9949**	0,3870	**0,9819**	0,5577	**0,9965**	0,9412
1%	3%	0,9660	**0,9996**	**0,8986**	0,1002	**0,9311**	0,1821	**0,9872**	0,9137
1%	5%	0,9608	**0,9974**	**0,7794**	0,1000	**0,8607**	0,1818	**0,9758**	0,9137
3%	1%	0,9892	**0,9932**	0,9993	**0,9964**	0,9942	**0,9948**	0,9972	**0,9975**
3%	3%	0,9886	**0,9989**	**0,9338**	0,5590	**0,9604**	0,7168	**0,9814**	0,8934
3%	5%	**0,9867**	0,9863	**0,9365**	0,4612	**0,9609**	0,6285	**0,9816**	0,8684
5%	1%	0,9929	**0,9942**	**0,9995**	0,9989	0,9962	**0,9966**	0,9974	**0,9976**
5%	3%	0,9928	**0,9985**	**0,9751**	0,8912	**0,9839**	0,9418	**0,9889**	0,9618
5%	5%	0,9915	**0,9934**	**0,9748**	0,8516	**0,9831**	0,9171	**0,9884**	0,9466

Çalışma kapsamında atakların birlikte uygulandığında etkisini değerlendirme amacı ile yukarıda bahsedilen üç atak eşit oranlarda kullanılarak elde edilen karışık atak itme ve çekme türleri için veri kümesine enjekte edilmiştir. İtme türü için karışık ataklar için performans metrikleri sonuçları Tablo 5.7'de verilmiştir. Kesinlik metriği için yine T-DVM modeli ile daha yüksek oranlar elde edilmiştir. Fakat diğer atak modelleri ile kıyaslandığında iki model atasındaki fark daha azdır. Duyarlılık metriğinde %100'e ulaşan başarılar her iki model ile elde edilmiştir. Bu metrik için ODOT-DVM ile yaklaşık olarak %22 daha iyi sonuçlar elde edilmiştir. F1-skoru için sonuçlar değerlendirildiğinde, ODBOT-DVM ile %19 daha iyi sonuçlar elde edilmiştir. İki deney kümesi için T-DVM ile daha iyi sonuçlar elde edilirken geri kalanlarda ODBOT-DVM ile daha iyi sonuçlar elde edilmiştir. Doğruluk metriği için de aynı yorumları yapmak mümkündür. Fakat burada iyileşme oranı yaklaşık %2,5 ile sınırlı kalmıştır.

Çekme türü karışık atak türü performans metrikleri sonuçları Tablo 5.8'de sunulmuştur. Göreceli olarak itme türü atak modeline göre sonuçlarda bir düşüşten bahsetmek mümkündür. Kesinlik metriği için tüm deney kümelerinde T-DVM modeli daha iyi sonuçlar üretirken, duyarlılık metriğinde tüm deney kümelerinde ODBOT-DVM ile daha iyi sonuçlar elde edilmiştir. F1-skor ve doğruluk metriklerinde bir deney kümesinde T-DVM yöntemi ile saha iyi sonuçlar elde edilmiştir. Her iki metrik için de geri kalan sekiz deney kümesinde ODBOT-DVM ile daha iyi sonuçlar elde edilmiştir.

Tablo 5.7. *İtme türü karışık atak için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skoru		Doğruluk	
AO	DO	KA_İtme ODBOT	KA_İtme	KA_İtme ODBOT	KA_İtme	KA_İtme ODBOT	KA_İtme	KA_İtme ODBOT	KA_İtme
1%	1%	0,9929	**0,9995**	**0,9994**	0,9630	**0,9961**	0,9809	**0,9993**	0,9964
1%	3%	0,9918	**1,0000**	**0,9837**	0,1200	**0,9877**	0,2143	**0,9977**	0,9156
1%	5%	0,9897	**0,9992**	**0,9401**	0,1050	**0,9643**	0,1900	**0,9933**	0,9142
3%	1%	0,9974	**0,9986**	**0,9998**	0,9991	0,9986	**0,9988**	0,9993	**0,9994**
3%	3%	0,9957	**0,9996**	**0,9924**	0,9340	**0,9940**	0,9657	**0,9971**	0,9840
3%	5%	0,9931	**0,9957**	**0,9891**	0,8127	**0,9911**	0,8950	**0,9957**	0,9540
5%	1%	0,9979	**0,9987**	**1,0000**	1,0000	0,9989	**0,9994**	0,9993	**0,9996**
5%	3%	0,9964	**0,9994**	**0,9977**	0,9909	**0,9970**	0,9952	**0,9979**	0,9967
5%	5%	0,9944	**0,9961**	**0,9971**	0,9779	**0,9957**	0,9869	**0,9970**	0,9910

Tablo 5.8. *Çekme türü karışık ataklar için kesinlik, duyarlılık, F1-skoru ve doğruluk sonuçları*

		Kesinlik		Duyarlılık		F1-Skoru		Doğruluk	
AO	DO	KA_Çekme ODBOT	KA_Çekme	KA_Çekme ODBOT	KA_Çekme	KA_Çekme ODBOT	KA_Çekme	KA_Çekme ODBOT	KA_Çekme
1%	1%	0,9694	**0,9983**	**0,9899**	0,2659	**0,9795**	0,4200	**0,9960**	0,9296
1%	3%	0,9642	**0,9998**	**0,8509**	0,1005	**0,9040**	0,1827	**0,9827**	0,9138
1%	5%	0,9577	**0,9964**	**0,7077**	0,1004	**0,8139**	0,1824	**0,9690**	0,9137
3%	1%	0,9891	**0,9935**	**0,9985**	0,9922	**0,9938**	0,9928	**0,9970**	0,9965
3%	3%	0,9878	**0,9979**	**0,9190**	0,5016	**0,9522**	0,6677	**0,9777**	0,8795
3%	5%	0,9856	**0,9843**	**0,9360**	0,4696	**0,9601**	0,6359	**0,9812**	0,8702
5%	1%	0,9928	**0,9943**	**0,9997**	0,9990	0,9962	**0,9966**	0,9974	**0,9977**
5%	3%	0,9920	**0,9981**	**0,9697**	0,8673	**0,9807**	0,9281	**0,9868**	0,9534
5%	5%	0,9907	**0,9932**	**0,9768**	0,8425	**0,9837**	0,9116	**0,9888**	0,9434

Karışık atak tipi de dahil olmak üzere dört farklı atak veri kümesine enjekte edilerek iki farklı DVM modeli ile tespit edilmesine yönelik sonuçlar önceki tablolarda sunulmuştur. Sonraki iki tabloda ise modellerin atak modellerinde elde edilen sonuçları karşılaştırmalı olarak verilmiştir. Tablolarda hangi atak modelinde daha iyi souç üretildiyse kalın yazı tipi ile belirtilmiştir.

Tablo 5.9'da dört farklı atak modelinin performans metrikleri sonuçlarının ortalamaları verilmiştir. Kesinlik metriği için bakıldığında, rastgele ataklarda ODBOT-DVM ile en iyi sonuçlar elde edilirken, T-DVM modeli ile karışık atak modelinde en iyi

sonuçlar elde edilmiştir. Diğer metrik türleri için ortalama atak tipinde daha iyi sonuçlar elde edilmiştir

Tablo 5.9. *İtme türü ataklar için farklı atak türlerinde elde edilen sonuçların ortalamaları*

	Kesinlik		Duyarlılık		F1-Skoru		Doğruluk	
Atak Tipi	ODBOT-DVM	T-DVM	ODBOT-DVM	T-DVM	ODBOT-DVM	T-DVM	ODBOT-DVM	T-DVM
Rastgele	**0,9954**	0,9982	0,9882	0,7321	0,9917	0,7805	0,9976	0,9666
Ortalama	0,9947	0,9976	**0,9916**	**0,7746**	**0,9931**	**0,8089**	**0,9980**	**0,9736**
M-Ortalama	0,9820	0,9955	0,9435	0,5939	0,9614	0,6797	0,9883	0,9371
Karışık	0,9944	**0,9985**	0,9888	0,7670	0,9915	0,8029	0,9974	0,9723

Çekme türü atak tipleri için performans metriklerinin sonuçları Tablo 5.10'da verilmiştir. Tüm atak tipleri ve türleri için modifiye edilmiş ortalama atak modelinde en iyi sonuçlar elde edilmiştir. ODBOT-DVM ile doğruluk oranlarında ortalama, modifiyeli ortalama ve karışık atak tipleri için aynı sonuçlar elde edilmiştir.

Tablo 5.10. *Çekme türü ataklar için farklı atak türlerinde elde edilen sonuçların ortalamaları*

	Kesinlik		Duyarlılık		F1-Skoru		Doğruluk	
Atak Tipi	ODBOT-DVM	T-DVM	ODBOT-DVM	T-DVM	ODBOT-DVM	T-DVM	ODBOT-DVM	T-DVM
Rastgele	0,9808	0,9945	0,9179	0,5308	0,9455	0,6165	0,9854	0,9249
Ortalama	0,9810	0,9927	0,9276	0,5869	0,9516	0,6753	**0,9863**	0,6753
M-Ortalama	**0,9820**	**0,9955**	**0,9435**	**0,5939**	**0,9614**	**0,6797**	**0,9883**	**0,9371**
Karışık	0,9810	0,9951	0,9276	0,5710	0,9516	0,6575	**0,9863**	0,9331

Çalışma kapsamında 4 farklı atak tipi MovieLens100K veri setine enjekte edilerek daha önceki çalışmalarda tespit edilmiş ataklar, gelecekteki gürbüzlüğü sağlamak amacı ile DVM ile tespit edilmiştir. Öneri sistemlerinde kullanılan veri kümeleri ürün sayısının fazlalığı dolayısıyla seyrektir. Bu veri kümesine eklenen ataklar ile gerçek kullanıcılar ile aradaki sayı farkından dolayı temel sınıflandırma yöntemleri yüksek performans oranlarına ulaşamamaktadır. Benzer tip ataklar sisteme enjektesini bir an evvel tespit edebilmek için aşırı örnekleme tabanlı DVM yöntemi önerilmiştir. Sınıflar arası eleman sayısı fazla olduğu için aşırı örnekleme ile performans metriklerinden elde edilen sonuçların geliştirilmesi sağlanmıştır. Bu doğrultuda aşırı örnekleme tabanlı DVM yöntemi ile şilin atakların tespiti yüksek doğruluk oranlarında sağlanabilmektedir.

6. ŞİLİN ATAKLARININ TEK SINIFLI DESTEK VEKTÖR MAKİNELERİ İLE TESPİTİ

Öneri sistemleri çeşitli çevrimiçi platformlarda hayati bir rol oynar ve kullanıcıların tercihlerini göz önünde bulundurarak yeni ürünler, hizmetler ve içerikler keşfetmelerine yardımcı olur. Bununla birlikte, bu sistemler, kötü niyetli kullanıcıların derecelendirmeleri yapay olarak şişirdiği veya söndürdüğü ve önyargılı önerilere yol açtığı şilin saldırıları yoluyla manipülasyona karşı savunmasızdır. Bu saldırıları araştırmanın, anlamanın ve hafifletmenin önemini vurgulamak çok önemlidir. Bu tür saldırıları tespit etmek, tavsiye sistemlerinin bütünlüğünü ve etkinliğini korumak için çok önemlidir. Literatürde, şilin saldırılarını tespit etmek için birçok çalışma sunulmuştur. En iyi bilinen kümeleme yöntemleri farklı saldırı modelleri için uyarlanmıştır. Bu makalede, şilin saldırılarını tespit etmek için gürbüz bir teknik olarak Tek Sınıflı Destek Vektör Makineleri (TSDVM) kullanımını araştırıyoruz. TSDVM, öncelikle anomali tespiti ve aykırılık tespiti görevleri için tasarlanmış geleneksel Destek Vektör Makinelerinin (DVM) özel bir çeşididir. Önerilen yöntemi doğrulamak için MovieLens100K veri kümesi kullanılmıştır. Sonuç olarak, farklı boyut ve doluluk oranlı saldırılar için hassasiyet ve geri çağırma değerleri verilmiştir.

6.1. TSDVM ve Şilin Ataklar

Öneri sistemleri, yeni ürünler, hizmetler ve içerikler keşfetmemize yardımcı olarak dijital hayatımızın vazgeçilmezi haline gelmiştir. Bu sistemler, kişiselleştirilmiş öneriler sunmak için kullanıcı tercihlerinden ve davranışlarından yararlanır. Öneri sistemleri gerçek dünyada e-ticaret (Frey vd., 2016; Ge vd., 2020), sağlık (Han vd., 2020; Yang & Jiang, 2018), enerji (alsalemi vd., 2020; Sardianos vd., 2020), e-öğrenme (Kulkarni vd., 2020; Li & Kim, 2021) ve sosyal ağlar (Puglisi vd., 2015) gibi geniş uygulama alanlarına sahiptir.

Artan popülerliklerine rağmen, öneri sistemleri şilin saldırıları olarak bilinen kalıcı bir tehditle karşı karşıyadır (Bilge vd., 2014; Rezaimehr & Dadkhah, 2020; Si & Li, 2020b). Bu saldırılar, bu sistemlerin güvenilirliğini ve bütünlüğünü zayıflatarak potansiyel olarak önyargılı önerilere ve tehlikeye atılmış kullanıcı deneyimlerine yol açmaktadır. Bu konuya ışık tutmak ve öneri sistemlerini bu tür saldırılardan korumak için potansiyel çözümleri araştırmak çok önemlidir. Bu saldırıları tespit etmek için sınıflandırma (Batmaz vd., 2020; Zhang & Zhou, 2014), kümeleme (Bhaumik vd., 2011;

Zahra vd., 2015; Zhang & Wang, 2020) ve istatistik (Bhaumik vd., 2006) tabanlı yaklaşımlar önerilmiştir.

Şilin atakları, sistem tarafından üretilen sonuçları değiştirmek amacıyla öneri sisteminin algoritmalarına yanlış veya taraflı veriler eklenmesini gerektirir. Kötü niyetli kullanıcılar, önerileri kendi lehlerine etkilemek için çeşitli hayali kullanıcı kimlikleri oluşturabilir, ürünleri dürüst olmayan bir şekilde inceleyebilir veya kurnazca yorumlar sunabilir. Saldırganlar, algoritmaların kullanıcı tercihleri hakkındaki bilgisini değiştirmeye ve sistemin geçmiş kullanıcı verilerine olan güveninden yararlanarak önerileri çarpıtmaya çalışırlar.

Şilin atakları yalnızca kullanıcı güvenini zedelemekle kalmaz, aynı zamanda öneri sistemlerinin bütünlüğünü ve adilliğini de tehlikeye atar. Saldırganlar temel algoritmaları değiştirebilir ve yanlış veya manipüle edilmiş veriler enjekte ederek öneri sürecine zarar verebilir. Bu, belirli şeyleri veya tercihleri korurken diğerlerini marjinalleştiren basamaklı bir etkiye sahip olabilir. Sonuç olarak, farklı ve faydalı öneriler bastırılabilir, bu da kullanıcı deneyimlerinin sınırlanmasına ve yeni bilgilere sınırlı maruz kalınmasına neden olur.

Tespit olasılığını azaltmak için farklı saldırı modelleri oluşturulur. Temel atak modelleri rastgele ve ortalama olup literatürdeki en basit saldırı modelleridir. Bu atak modellerinde seçilmiş öğe kısımları yoktur. Rastgele saldırı modelinde, dolgu ürünleri genel ortalamaya göre derecelendirilir (Lam & Riedl, 2004). Diğer yandan, ortalama saldırı modelinde dolgu ürünleri ürünlerin puan ortalamalarına göre derecelendirilir (Burke vd., 2005). Hedef ürün, atak türüne göre maksimum veya minimum puanla derecelendirilir. İtme saldırıları için, hedef ürün maksimum puanla derecelendirilir. Çekme atakları için, hedef ürün minimum puanla derecelendirilir. Sürü atakları rastgele ve ortalama saldırıların uzantısıdır. Literatürde iki tür sürü atağı vardır; rastgele ve ortalama sürü atakları. Her iki sürü atak modelinde de popüler ürünler seçilir ve hedef ürün ile maksimum puanlar verilir. Dolgu ürünleri rastgele ve ortalama ataktakine benzer şekilde derecelendirilir.

Bölüm atakları ürün tabanlı öneri sistemlerinde daha etkilidir. Bölüm ataklarında belirli bir ürünle ilişkili kullanıcılar seçilir ve bu kullanıcılara öneriler sunulur (Mobasher vd., 2005). Yoklayıcı ataklarında kullanıcı veri tabanından oluşturulan çekirdek profiller

kullanılmaktadır (Mobasher vd., 2007). Kullanılan profillerde hedef ürün puanı manipüle edilmektedir. Yukarıda bahsedilen atak modelleri genellikle itme saldırı modelleridir. Hedef ürünün popülerliğinin artmasını sağlarlar (O'Mahony vd., 2005). Hedef öğenin popülerliğini azaltmak için literatürde sevgi/nefret, ters sürü ve mükemmel bilgi atak modelleri önerilmiştir (Mobasher vd., 2007; Williams & Mobasher, 2006).

Literatürde şilin atak tespit yöntemleri üç ana alanda incelenmektedir: denetimsiz, yarı denetimli ve denetimli. Denetimsiz yöntemler diğer yöntemlere göre daha çok tercih edilmektedir. Farklı kümeleme yöntemleri ile mevcut veriler için herhangi bir sınıf etiketine ihtiyaç duyulmadan saldırılar tespit edilebilmektedir. Atak profilleri tarafından yapılan derecelendirmeler anormal bir dağılım sergileyeceğinden, gerçek profillerin oluşturduğu kümelerin dışında kalacaktır.

Literatürde şilin atak tespiti için farklı kümeleme yöntemleri kullanılmaktadır. Bu çalışmada, Markov zincirleri ile kullanıcıların şüpheli kullanıcı derecelendirmeleri çıkarılmaktadır. Daha sonra elde edilen şüphelilik derecelerinin hiyerarşik kümelenmesi ile şüpheli kullanıcılar tespit edilmektedir (Zhang vd., 2018). Bir başka çalışmada ise dört tespit özelliği çıkarılarak veri tabanındaki kullanıcıların puanlama davranışları incelenmiştir. Daha sonra PCA ve k-ortalamalar yöntemleri kullanılarak şüpheli kullanıcılardan oluşan bir küme oluşturulmaktadır. Şüpheli kullanıcıların toplanma davranışları saldırganlar ile gerçek kullanıcıları ayırt etmek için değerlendirilmektedir. Böylece doğruluk oranının artırılması hedeflenmektedir (Cai & Zhang, 2019).

Yazarlar, yumuşak ortak kümeleme algoritmasını kullanıcı eğilimi benzerliği yöntemiyle entegre etmenin avantajlarını araştırmış ve şilin ataklarını tespit etmek için eğilim benzerliği ile yumuşak bir ortak kümeleme modeli sunmuşlardır (Yang vd., 2017). Bir başka çalışma, bir profilin saldırgan olması için sahip olması gereken nitelikleri araştırmaktadır. Kullanıcı-ürün puan matrisi, kullanıcı-bağlantı matrisi ve kullanıcılar arasındaki benzerlik, derecelendirmeler arasında anormallik olup olmadığını belirlemek için kullanılır. Elde edilen bilgiler ile k-ortalamalar yöntemi saldırgan kullanıcıları tespit etmektedir (Davoudi & Chatterjee, 2017). Çalışmada atak kullanıcılar için yeni bir öznitelik önerilmiştir. Daha sonra şilin atakları için üç alternatif tespit modeli önerilmiştir (Panagiotakis vd., 2020).

Denetimli yöntemler, diğer verilere sınıflandırma yapmak için etiketli verilere ihtiyaç duyar. Öneri sistemlerinde, destek vektör makineleri atak kullanıcılarını tespit etmek için tercih edilir. (Alostad, 2019)'de saldırı tespiti için modifiye edilmiş bir SVM modeli kullanılmıştır. Yüksek tespit oranları elde etmek için makine öğrenimi modeli olarak Gauss Karışım Modeli kullanılmıştır. Bazı senaryolarda, tespit için az sayıda etiketli veri vardır. Bu durumda, yarı denetimli yöntemler şilin ataklarının tespit edilmesine yardımcı olur. İyi bilinen sınıflandırma yöntemi Naïve Base, şilin ataklarını sınıflandırmak için kullanılır (Zhang vd., 2017). Bu çalışma, başlangıç sınıflandırıcısını iteratif olarak geliştirmek için Beklenti Maksimizasyonu (EM) algoritmasını kullanmaktadır. Başka bir çalışmada (Zhou & Duan, 2021), yarı denetimli atak tespiti için Co-Forest algoritması kullanılmıştır.

DVM, sınıflandırma problemleri için güçlü bir yöntemdir. Ancak, verileri sınıflandırmak için verileri etiketlemesi gerekir. Bu çalışmada, verilerdeki şilin atakları tespit etmek için TSDVM kullanılmıştır. TSDVM, makine öğrenimi ve veri analizindeki çeşitli zorlukların üstesinden gelmek için ilgi çekici ve güçlü bir aracı temsil eder. TSDVM, öncelikle anomali tespiti ve yenilik tespiti görevleri için tasarlanmış geleneksel Destek Vektör Makinelerinin özel bir çeşididir.

TSDVM uygulanması, bilgisayar bilimleri, mühendislik, sosyal bilimler ve hatta doğa bilimleri de dahil olmak üzere birçok alanda özellikle değerli olabilir. Bu algoritmalar, araştırmacıların veri noktalarının çoğundan farklı olan nadir veya olağandışı gözlemleri belirlemeye odaklandıkları problemlerin üstesinden gelmelerine olanak tanır. Bunu yaparak, TSDVM istisnai kalıpları anlamaya, aykırı değerleri tespit etmeye ve normdan önemli ölçüde sapan anomalileri karakterize etmeye yardımcı olur.

TSDVM yüksek boyutlu bir özellik uzayını iki bölgeye ayırma prensibiyle çalışır: verilerin çoğunluğunu içeren bölge (normal örnekler) ve aykırı değerleri veya anomalileri temsil eden bölge. Bu ayırma işlemi, normal örnekler ile özellik uzayının orijini arasındaki marjı maksimize eden optimal bir hiperdüzlem oluşturularak gerçekleştirilir. Bu şekilde, TSDVM normal veri örneklerinin özelliklerini kapsüllemeyi öğrenir ve bunların ayrılabilirliğinin bir ölçüsünü sağlar.

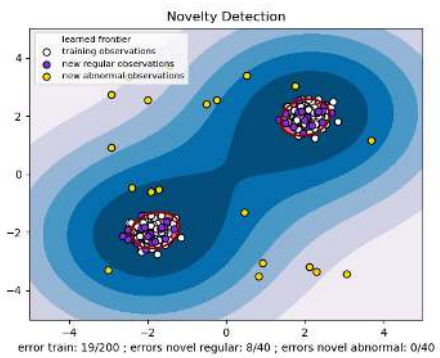
TSDVM çeşitli senaryolarda kullanılabilir. Örneğin, bilgisayar bilimlerinde, bu algoritmalar siber güvenlik tehditlerini tanımlamaya, ağ izinsiz girişlerini tespit etmeye

veya kötü amaçlı yazılımları iyi huylu programlardan ayırt etmeye yardımcı olabilir (Wang vd., 2023). Mühendislikte, TSDVM arıza teşhisine, endüstriyel süreçlerde anormallik tespitine veya üretimde kalite kontrolüne yardımcı olabilir (Cao vd., 2020; Li vd., 2020). Sosyal bilimler, insan faaliyetlerindeki anormal davranış kalıplarını, finanstaki hileli işlemleri veya sosyolojik çalışmalardaki anormal eğilimleri tanımlamak için TSDVM'den yararlanabilir (Abdulhussein & Nasrudin, 2022; Karasmanoglou, 2023). Doğa bilimlerinde bile bu algoritmalar nadir olayların, çevresel verilerdeki aykırı değerlerin veya biyolojik sistemlerdeki anormalliklerin tespit edilmesine yardımcı olabilir (Cheng vd., 2019).

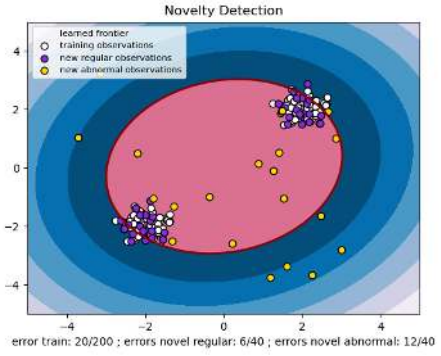
TSDVM en önemli avantajlarından biri, sınırlı eğitim verileriyle çalışabilmeleridir. Anomali tespit problemlerinin tipik olarak seyrek ve dengesiz veri kümeleri içerdiği göz önüne alındığında, TSDVM aykırı değerlerin varlığına karşı dayanıklı olurken normal örneklerin altında yatan yapıyı yakalamada mükemmeldir. Dahası, bu algoritmalar sınıflandırmayla ilişkili bir güven veya olasılık ölçüsü sağlayarak araştırmacılara tespit edilen anomalilerin önemini yorumlamada yardımcı olur.

Sayısız avantajlarına rağmen, TSDVM'nin de belirli sınırlamaları vardır. Verileri yüksek boyutlu uzaylara eşlemekten sorumlu olan uygun çekirdek fonksiyonlarının seçimi, bu algoritmaların performansını önemli ölçüde etkileyebilir. Ayrıca, çekirdek genişliği ve düzenleme parametresi gibi optimum parametrelerin seçilmesi genellikle dikkatli ayarlama ve çapraz doğrulama gerektirir (Anaissi vd., 2018; Ghafoori vd., 2018; Guerbai vd., 2021; S. Wang vd., 2017, 2018; Xiao vd., 2015). TSDVM için v ve ρ olmak üzere iki parametre vardır. v, 0 ile 1 arasında değerler alır ve gevşeme seviyesini kontrol eder ve ρ ise yanlılık terimidir. Bu parametreler bir sonraki bölümde ayrıntılı olarak açıklanmaktadır. Bu parametrelerin seviyesi yöntemin doğruluk oranı üzerinde oldukça etkilidir. ***Şekil 6.1***'de farklı parametre seviyelerinin etkisi verilmiştir. ***Şekil 6.1.a***'da parametrelerin temel seviyeleri verilmiştir. Bu seviyeler için eğitim kümesi hatası %9,5, normal veri için hata %20 ve anormal veri için hata %0'dır. ***Şekil 6.1.a***'dan da görebileceğimiz gibi, bu senaryo için bir aşırı öğrenme gözlemlenmektedir. Tüm aykırı değerler doğru sınıflandırılmıştır, ancak normal öğe için hatalar yüksektir. v aynı iken ρ azaltılırsa sınırlar çok büyük olur ve anormal öğeler normal bölgede yer alır. İki parametre seviyesi azaltıldığında, normal öğelerin hataları düşüktür, ancak anormal öğelerin hatası hala yüksektir. Parametre seviyeleri $v = 0{,}01$, $\rho = 0{,}1$ olarak seçilirse, en iyi doğruluk

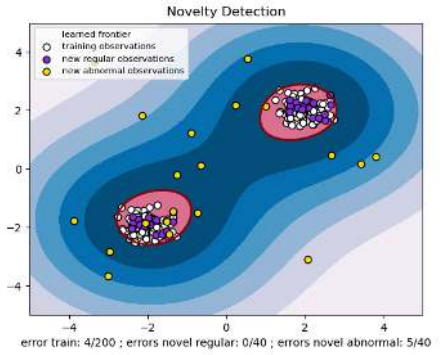
oranı elde edilir. Ancak, bu seviye optimal değildir. Parametre seçimi için farklı yöntemler önerilmiştir, bu yöntemler bu bölümün sonunda verilecektir.



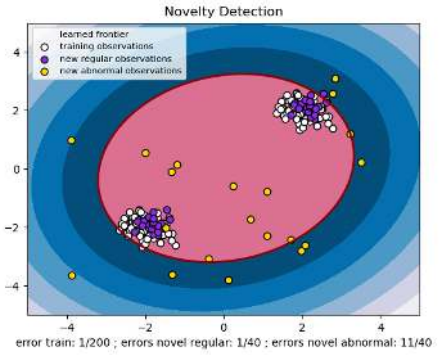
Şekil 6.1.a. *v*=0.1 *ρ*=0.1



Şekil 6.1.b. *v=0.1 ρ=0.01*



Şekil 6.1.c. *v=0.01 ρ=0.1*



Şekil 6.1.d. *v=0.01 ρ=0.01*

Şekil 6.1. *Farklı hiperparametre değerlerinin TSDVM sonuçları üzerine etkisi*

Bu çalışmada, TSDVM yöntemi şilin saldırı tespiti için öneri sistemine uyarlanmıştır. TSDVM'nin hiperparametreleri denetimsiz bir yöntemle seçilmiştir.

6.2. Tek Sınıflı Destek Vektör Makineleri

OCSVM sınıflandırıcısını tanımlayan primal kuadratik problem şöyledir

$$\begin{aligned} &\min_{\omega,\xi,\rho} \frac{1}{2}\left\|\omega^2\right\| + \frac{1}{vn}\sum_{i=1}^{n}\xi_i - \rho \\ &s.t.\left\langle \omega.\varphi\left(x_i\right)\right\rangle \geq \rho - \xi_i \\ &\xi_i \geq 0, i = 1,2,...,n \end{aligned} \tag{29}$$

Burada $\omega \in R^d$ ve $\xi = [\xi_1,...,\xi_n]$ gevşek değişkenlerdir. ρ yanlı terim ve v $0 < v \leq 1$ aralığında değerler alır. Gevşek değişkenler, problemin dayattığı

kısıtlamaların gevşetilmesini sağlar. v parametresi gevşeme seviyesini kontrol eder. Eğer 0'a yakın ayarlanırsa, ceza terimi $\sum_{i=1}^{n}\xi_i$ yok olur. Başka bir deyişle, aşırı öğrenmeye neden olur, model veri kümesindeki tüm noktaları ayırmayı amaçlar. Buna karşılık, bu parametre 1'e çok yakınsa, algoritma (29)'u en aza indirmek için daha fazla serbestliğe sahiptir. Bu durumda yetersiz uyum gözlenir. Noktaların çoğu anomali olarak etiketlenir.

(1), hesaplamaktan kaçınmak için Lagrange çarpanları kullanılarak aşağıdaki ikinci dereceden dual probleme dönüştürülür:

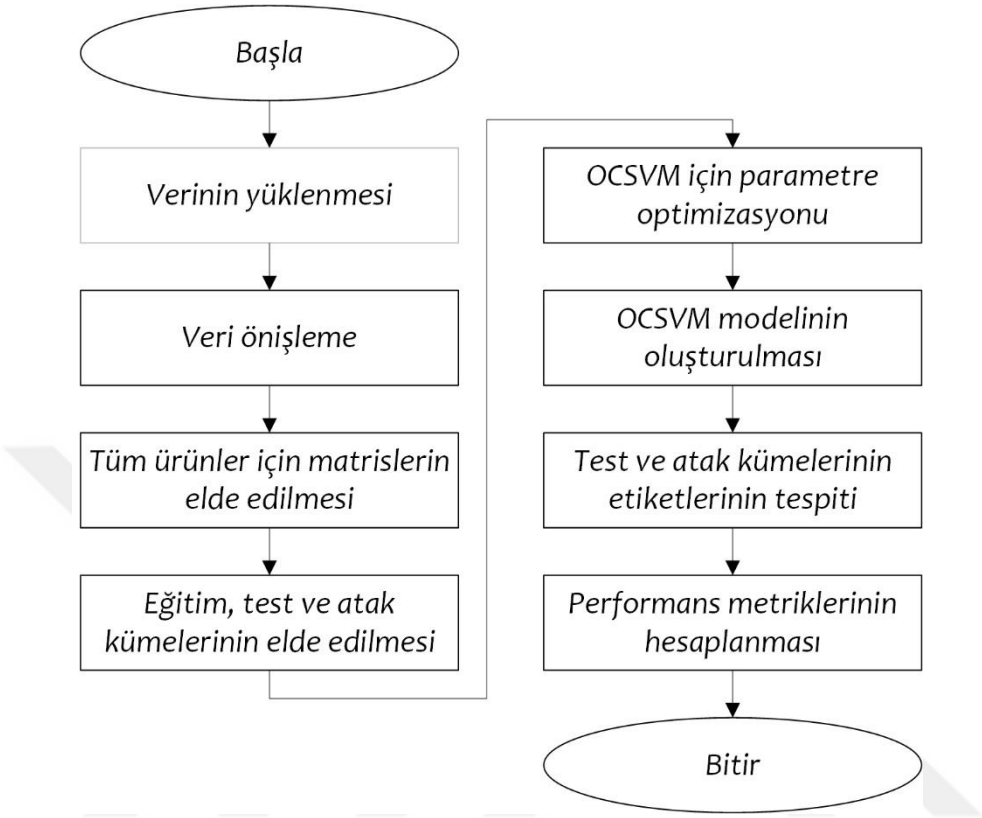
$$\begin{aligned} &\min_{[a_i]} \frac{1}{2}\sum_{i,j} a_i a_j k\left(x_i, x_j\right) \\ &s.t. 0 \le a_i \le 1/vn \\ &\sum_{i=1}^{n} a_i = 1,\ i = 1,2,...,n \end{aligned} \tag{30}$$

Şilin atakları, kullanıcıların öneri sistemlerine olan güveni için önemli bir tehdit oluşturmaktadır. Kullanıcılar önerileri taraflı veya manipüle edilmiş olarak algıladıklarında sisteme olan güvenleri azalır. Kullanıcılar önerilen öğelerin gerçekliğini ve sistemin tercihlerini doğru bir şekilde anlama becerisini sorgulayabilir. Bu güven erozyonu, katılımın azalmasına, memnuniyetin düşmesine ve hatta öneri sisteminin tamamen terk edilmesine yol açabilir.

Bu çalışmada, MovieLens100K veri kümesi açıklayıcı bir örnek olarak kullanılmıştır. Bu veri kümesi 943 kullanıcının 1652 filme verdiği 100.000 puanı içermektedir. Puanlar 1 ile 5 arasında tam sayı değerlerini almaktadır. Bu çalışmada, model doğrulama için k-kat çapraz doğrulama uygulanmıştır. Eğitim veri kümesinde sadece gerçek profiller kullanılmıştır. Eğitim veri kümesi için gerçek profillerin %80'i, test veri kümesi için ise gerçek profillerin kalan %20'si kullanılmıştır. Ek olarak, saldırı profilleri oluşturulan model ile test edilmiştir. Her replikasyon için farklı parçalar eğitim ve test örnekleri olarak rastgele seçilmiştir.

Bu çalışmada, verilere rastgele, ortalama ve karışık ataklar olmak üzere üç tür enjekte edilmiştir. Karma atak rastgele ve ortalama ataklardan oluşmaktadır. Bu saldırılar veri kümesindeki en çok puan alan 50 ürüne enjekte edilmiştir. Her bir ürün için 10

tekrarlama sağlanmıştır. Çalışama da uygulanan adımların akış şeması Şekil 6.2'de verilmiştir.



Şekil 6.2. *Tek sınıflı destek vektör makinaları ile atak tespiti akış şeması*

Bu çalışmada uygulanan yöntemin algoritması **Algoritma 6.1**'de verilmiştir.

Algoritma 6.1: TSDVM ile şilin atak tespiti

Girdi: Kullanıcı-Ürün Matrisi (X)

for $k \leftarrow 1$ to 50 **do**
    50 popüler ürünü bul
    Her ürün için kullancı-ürün matrisini çıkart
    Öğrenme, test ve atak kümelerini oluştur
    TSDVM için optimal parametre değerlerini bul
    for $j \leftarrow 1$ to 5 **do**
        i. ürün j. öğrenme kümesi için TSDVM modelini kur
        j. öğrenem kümesi için etiketleri tespit et
        Doğruluk oranlarını hesapla
        for $k \leftarrow 1$ to 10 do
            i. ürüne enjecte edilen k. atak kümesi için etiketleri tespit et
        end for
        Duyarlılık ve kesinlik değerlerini hesapla
    end for
end for

Öğrenme kümesi için kurulan modellerin doğruluk oranları Tablo 6.1'de verilmiştir. Uygulamada kullanılan k-kat çapraz doğrulamadan dolayı öğrenme kümesi doğruluk oranlarında farklılıklar gözlemlenmektedir.

Tablo 6.1. *Tek sınıflı destek vektör makinaları öğrenme kümesi doğruluk sonuçları*

	Atak Oranı	1			3			5		
	Dolgu Oranı	1	3	5	1	3	5	1	3	5
Doğruluk	Rastgele	0,9795	0,9806	0,9795	0,9812	0,9805	0,9792	0,9821	0,9821	0,9804
	Ortalama	0,9800	0,9798	0,9795	0,9813	0,9806	0,9800	0,9827	0,9810	0,9807
	Karma	0,9793	0,9802	0,9795	0,9821	0,9801	0,9795	0,9825	0,9798	0,9800

Çalışmanın tüm test sonuçları için Tablo 6.2 elde edilmiştir. Tablo 6.2'de tüm saldırı modelleri için farklı atak ve dolgu oranı için duyarlılık, kesinlik ve F1-skoru değerleri verilmiştir. Genel olarak F1-skoru için yüzde 92'nin üzerinde sonuçlar elde edilmiştir. Özellikle düşük dolgu oranları yüksek dolgu oranlarına göre daha iyi sonuçlar vermektedir. Bu çalışmada TSDVM'nin hiperparametrelerini tahmin etmek için hızlı model seçimi (Quick Model Selection) (Ghafoori vd., 2018) yöntemi kullanılmıştır. Farklı hiperparametre tahmin yöntemleri, performans metrikleri için daha yüksek puanlar elde edebilir.

Tablo 6.2. *MovieLens100K veri kümesi için TSDVM ile atak tespiti performans ölçütleri sonuçları*

	Atak .	1			3			5		
	Dolgu	1	3	5	1	3	5	1	3	5
Duyarlılık	Rastgele	1,0000	0,9994	0,9946	1,0000	0,9991	0,9840	1,0000	0,9982	0,9741
	Ortalama	1,0000	0,9997	0,9972	1,0000	0,9996	0,9904	1,0000	0,9990	0,9843
	Karma	1,0000	0,9999	0,9974	1,0000	0,9995	0,9896	1,0000	0,9989	0,9849
Kesinlik	Rastgele	0,9431	0,9424	0,9423	0,9479	0,9445	0,9486	0,9486	0,9475	0,9430
	Ortalama	0,9465	0,9431	0,9424	0,9493	0,9458	0,9510	0,9510	0,9479	0,9457
	Karma	0,9465	0,9431	0,9424	0,9493	0,9458	0,9510	0,9510	0,9479	0,9457
F1-Skoru	Rastgele	0,9707	0,9701	0,9678	0,9733	0,9710	0,9625	0,9736	0,9722	0,9583
	Ortalama	0,9725	0,9706	0,9690	0,9740	0,9720	0,9664	0,9749	0,9728	0,9646
	Karma	0,9720	0,9705	0,9691	0,9734	0,9714	0,9658	0,9749	0,9725	0,9641

Bu çalışmada, gerçek kullanıcılar pozitif gözlemler olarak kabul edilmiştir. Buna karşılık, saldırı kullanıcıları pozitif gözlem olarak kabul edilirse Tablo 6.2'de verilen sonuçlar azalabilir. Ancak, ele alınan veri seyrek bir veridir. Verilerin doldurulma oranları yaklaşık %6'dır. Bu veri için uygun bir hiperparametre seçim yöntemi seçilirse, sonuçlar her iki durum için de daha iyi olacaktır.

6.3. TSDVM ile Atak Tespiti Sonuçları

Öneri sistemleri, aşırı bilgi yükü sorunlarıyla başa çıkmak için güçlü bir tekniktir. Bununla birlikte, şilin atakları öneri sistemlerinin güvenilirliği ve bütünlüğü için önemli bir zorluk teşkil etmektedir. Bu atakları araştırmanın, anlamanın ve hafifletmenin önemini vurgulamak çok önemlidir. Sağlam savunmalar geliştirerek ve alandaki bilgi birikimini ilerleterek daha güvenilir, adil ve kullanıcı merkezli öneri sistemlerini teşvik edebiliriz. Nihayetinde bu, kullanıcı memnuniyetinin artmasına, güvenin yükselmesine ve bireysel tercihler ve ilgi alanlarıyla uyumlu daha anlamlı önerilerin verilmesine yol açacaktır.

Literatürde, şilin ataklarını tespit etmek için denetimli, yarı denetimli ve denetimsiz yöntemler önerilmiştir. Çalışmada, matematiksel bir tespit algoritması öneri sistemine uyarlanarak şilin atakları tespiti yapılmaktadır. TSDVM'nin araştırma ve analize dahil edilmesi, çeşitli alanlardaki anormalliklerin anlaşılmasını ve tespit edilmesini büyük ölçüde geliştirebilir. Nadir olayları, aykırı değerleri ve olağandışı örüntüleri belirleme ve karakterize etme yetenekleri, değerli içgörüler sağlar ve bilgisayar bilimi ve mühendisliğinden sosyal bilimlere ve doğa bilimlerine kadar çeşitli alanlarda ilerlemelere katkıda bulunabilir. TSDVM'nin gücünden yararlanılarak anomali tespitinin ortaya çıkardığı zorlukları etkili bir şekilde ele alınabilirler. Farklı hiperparametre seçim yöntemleri araştırılmış ve TSDVM hiperparametreleri için denetimsiz bir yöntem seçilmiştir. Bu çalışmada TSDVM yöntemi MovieLens100K veri kümesi için farklı saldırı ve dolgu boyutları için uygulanmıştır. Tüm senaryolar için doğruluk oranları verilmiştir.

7. SONUÇLAR

Öneri sistemleri, aşırı bilgi yükleme sorununun ele alınmasında önemli bir rol oynamaktadır. Bununla birlikte, öneri sistemleri, şilin atakları veya profil enjeksiyon ataklarına karşı savunmasızdır. Bu çalışmada, yeni bir çevrimiçi şilin atağı tespit yöntemi önerilmiştir. Literatürde sıklıkla kullanılan kalite kontrol grafiği ÜAHO ve popüler alternatiflerine göre daha az bilinen SR prosedürü ile şüpheli derecelendirmelerinin tespiti için kullanılmıştır. Yüksek duyarlılık oranı için, kesinlik oranından ödün verilerek tüm atakların tespiti hedeflenmiştir. Şüpheli puanlamalar veren bu şüpheli kullanıcılar, uzaklık metriği olarak korelasyonun kullanıldığı yoğunluk tabanlı kümeleme yöntemi (DBSCAN) kullanılarak kümelenmiştir. Var olan yöntemlerin aksine tüm veri setinin kümelenmesindense sadece şüpheli kullanıcıların tespiti üzerinde durulmuştur. Bahsedilen çalışmalardan farklı olarak önerilen yöntem araştırılan kullanıcı sayısını azaltmaktadır. Bu sayede önerilen yöntemde yüksek duyarlılık ve kesinlik değerleri elde edilmektedir.

Atak profilleri düşük korelasyona veya benzerliğe sahiptir. Hedef ürün harici yapılan puanlamalar ve atak zamanı, atak kullanıcıları arasında çok benzerdir. Literatürdeki çalışmaların çoğu atakların zamanını dikkate almamaktadır. Kümeleme veya sınıflandırma yaklaşımları, kullanıcılar arasındaki mesafeyle sınırlıdır. Bu çalışmada, yüksek duyarlılık ve kesinlik oranları elde etmek için başka bir tespit özelliği olarak atak zamanı kullanılmıştır. Farklı ardışık atak türleri ayrıntılı olarak incelenmiştir.

Şüpheli profillerin tespiti için Shiryaev-Roberts prosedürü alternatif bir kalite kontrol grafiği olarak sunulmuştur. SR prosedürü ile puanlamalardaki ortalama sapmaları ve varyans üzerindeki değişimlerin tek grafikte saptanması amaçlanmıştır.

Kalite kontrol grafikler ile tespit edilen atak profilleri DVM ile eğitilerek gelecekte benzer saldırıların engellenmesi için kullanılmıştır. Bu sayede atak profillerin tespit edilmesine hız kazandırılmıştır. Burada ataklar ile gerçek kullanıcıların sayısı karşılaştırıldığında aradaki farkın kurulan DVM modelinin performansını düşürdüğü saptanmıştır. Bu yüzden model performansını artırmak amacıyla son zamanlarda önerilen ODBOT yöntemi kullanılarak sınıflar arasındaki sayı farkı azaltılmıştır. ODBOT kullanılan ve kullanılmayan modellerin karşılaştırmalı sonuçları sunulmuştur. İki modelin performans ölçütleri karşılaştırıldığında ODBOT kullanılan modelde ciddi oranda performans iyileşmesi gözlemlenmiştir.

Son olarak atakların tespitinde matematiksel yöntemlerden faydalanmak için TSDVM kullanılmıştır. Burada farklı hiperparametre seçim yöntemleri değerlendirilerek Quick Model Selection ile hiperparametre seçimi gerçekleştirilmiştir. Veri kümesinin farklı olmasından dolayı yüksek performans ölçüt değerleri elde edilmese de sonuçların geliştirilebilir olduğu gözlemlenmiştir.

Sonuç olarak bu tez kapsamında şilin ataklarında atağın yapıldığı zaman ayrı bir kriter olarak kullanılmış ve atağın tespitinde rol oynamıştır. Sistemin çevrimiçi tespitine olanak sağlanarak olası ataklara erken müdahale mümkün kılınmıştır. Şüpheli kullanıcıların tespiti aşamasında hızlı ve etkin sonuçlar üreten DBSCAN algoritması ile atak profiller gerçek profillerden ayırt edilmiştir. Tespit edilen atakların gelecekte sisteme eklenmesini erken tespit edebilmek amacı ile DVM modeli eğitilerek atakların tespitinin hızlandırılması hedeflenmiştir. Alternatif şilin atak tespit yöntemi olarak TSDVM ile atak tespitine yönelik çalışma gerçekleştirilmiştir.

Bu çalışmada kullanılan yöntemler genellikle farklı veri kümeleri için parametreler kullanmıştır. Gelecek çalışmalarda öneri sistemlerinde kullanılan veri kümeleri için hiperparametre seçimine yönelik yöntemler araştırılabilir.

KAYNAKÇA

Abdulhussein, A. A., & Nasrudin, M. faisun. (2022). A Genetic Algorithm Based One Class Support Vector Machine Model for Arabic Skilled Forgery Signature Verification. *SSRN Electronic Journal*.

Afsahhosseini, F., & Al-Mulla, Y. (2021). Smart, hybrid and context-aware POI mobile recommender system in tourism in Oman. *Journal of Cultural Heritage Management and Sustainable Development*, *ahead-of-p*(ahead-of-print).

Alostad, J. M. (2019). Improving the Shilling Attack Detection in Recommender Systems Using an SVM Gaussian Mixture Model. *Journal of Information and Knowledge Management*, *18*(1), 1–18.

alsalemi, abdullah, Himeur, Y., Bensaali, F., amira, abbes, Sardianos, C., Varlamis, I., & Dimitrakopoulos, G. (2020). achieving Domestic Energy Efficiency Using Micro-Moments and Intelligent Recommendations. *IEEE Access*, *8*, 15047–15055.

Althbiti, A., Alshamrani, R., Alghamdi, T., Lee, S., & Ma, X. (2021). Addressing Data Sparsity in Collaborative Filtering Based Recommender Systems Using Clustering and Artificial Neural Network. *2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021*, 218–227.

Anaissi, A., Braytee, A., & Naji, M. (2018). Gaussian Kernel Parameter Optimization in One-Class Support Vector Machines. *Proceedings of the International Joint Conference on Neural Networks*, *2018-July*(c).

Astorino, A., & Gaudioso, M. (2002). Polyhedral separability through successive LP. *Journal of Optimization Theory and Applications*, *112*(2), 265–293.

Atev, S., Masoud, O., & Papanikolopoulos, N. (2006). Learning traffic patterns at intersections by spectral clustering of motion trajectories. *IEEE International Conference on Intelligent Robots and Systems*, 4851–4856.

Bagirov, A. M. (2005). Max-min separability. *Optimization Methods and Software*, *20*(2–3), 271–290.

Barragáns-Martínez, A. B., Costa-Montenegro, E., Burguillo, J. C., Rey-López, M., Mikic-Fonte, F. A., & Peleteiro, A. (2010). A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. *Information Sciences*, *180*(22), 4290–4311.

Barros, R. C., Basgalupp, M. P., De Carvalho, A. C. P. L. F., & Freitas, A. A. (2012). A survey of evolutionary algorithms for decision-tree induction. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *42*(3), 291–312.

Batmaz, Z., Yilmazel, B., & Kaleli, C. (2020). Shilling attack detection in binary data: a classification approach. *Journal of Ambient Intelligence and Humanized Computing*, *11*(6), 2601–2611.

Bennett, K. P., & Mangasarian, O. L. (1992). Optimization Methods and Software Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, *1:1*, 23–34.

Bhaumik, R., Mobasher, B., & Burke, R. (2011). A Clustering Approach to Unsupervised

Attack Detection in Collaborative Recommender Systems. *Proceedings of the 7th IEEE International Conference on Data Mining*, 181–187.

Bhaumik, R., Williams, C., Mobasher, B., & Burke, R. (2006). Securing collaborative filtering against malicious attacks through anomaly detection. *AAAI Workshop - Technical Report*, *WS-06-10*(Shewart), 50–59.

Bilge, A., Ozdemir, Z., & Polat, H. (2014). A novel shilling attack detection method. *Procedia Computer Science*, *31*, 165–174.

Bilge, Alper, Ozdemir, Z., & Polat, H. (2014). A novel shilling attack detection method. *Procedia Computer Science*, *31*, 165–174.

Borràs, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, *41*(16), 7370–7389.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). Training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 144–152.

Bouyeddou, B., Harrou, F., Sun, Y., & Kadri, B. (2017). Detecting SYN flood attacks via statistical monitoring charts: A comparative study. *2017 5th International Conference on Electrical Engineering - Boumerdes, ICEE-B 2017*, *2017-Janua*, 1–6.

Burke, R., Mobasher, B., & Bhaumik, R. (2005). Limited Knowledge Shilling Attacks in Collaborative Filtering Systems 1 Introduction 2 Attack Models Knowledge and. *Information Systems*.

Burke, R., Mobasher, B., Bhaumik, R., & Williams, C. (2005). Segment-based injection attacks against collaborative filtering recommender systems. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 4–7.

Burke, R., Mobasher, B., Williams, C., & Bhaumik, R. (2006a). Classification features for attack detection in collaborative recommender systems. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2006*, 542–547.

Burke, R., Mobasher, B., Williams, C., & Bhaumik, R. (2006b). Detecting profile injection attacks in collaborative recommender systems. *CEC/EEE 2006 Joint Conferences*, *2006*.

Burke, R., Mobasher, B., Zabicki, R., & Bhaumik, R. (2005). Identifying Attack Models for Secure Recommendation. *Beyond Personalization*, 19.

Cai, H., & Zhang, F. (2019). An unsupervised method for detecting shilling attacks in recommender systems by mining item relationship and identifying target items. *Computer Journal*, *62*(4), 579–597.

Cai, H., & Zhang, F. (2021). BS-SC: An Unsupervised Approach for Detecting Shilling Profiles in Collaborative Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, *33*(4), 1375–1388.

Cao, C., Liu, M., Li, B., & Wang, Y. (2020). Mechanical fault diagnosis of high voltage circuit breakers utilizing VMD based on improved time segment energy entropy and

a new hybrid classifier. *IEEE Access*, *8*, 177767–177781.

Cao, J., Wu, Z., Mao, B., & Zhang, Y. (2013). Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system. *World Wide Web*, *16*(5–6), 729–748.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215.

Cevikalp, H., Uzun, B., Köpüklü, O., & Ozturk, G. (n.d.). *Deep Compact Polyhedral Conic Classifier for Open and Closed Set Recognition*.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2011). SMOTE: Synthetic Minority Over-sampling Technique. *Journal Of Artificial Intelligence Research*, *16*, 321–357.

Cheng, T., Dairi, A., Harrou, F., Sun, Y., & Leiknes, T. (2019). Monitoring influent conditions of wastewater treatment plants by nonlinear data-based techniques. *IEEE Access*, *7*, 108827–108837.

Cheng, Z., & Hurley, N. (2010a). Robust collaborative recommendation by least trimmed squares matrix factorization. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, *2*, 105–112.

Cheng, Z., & Hurley, N. (2010b). Robustness analysis of model-based collaborative filtering systems. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6206 LNAI*(May), 3–15.

Das, J., Mukherjee, P., Majumder, S., & Gupta, P. (2014). Clustering-based recommender system using principles of voting theory. *Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014*, 230–235.

Davoudi, A., & Chatterjee, M. (2017). Detection of profile injection attacks in social recommender systems using outlier analysis. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, *2018-Janua*, 2714–2719.

Frey, R. M., Vučkovac, D., & Ilic, A. (2016). A secure shopping experience based on blockchain and beacon technology. In G. I. & S. A. (Eds.), *CEUR Workshop Proceedings* (Vol. 1688). CEUR-WS.

Gao, M., Yuan, Q., Ling, B., & Xiong, Q. (2014). Detection of abnormal item based on time intervals for recommender systems. *The Scientific World Journal*, *2014*(2).

Gasimov, R. N., & Ozturk, G. (2006). Separation via polyhedral conic functions. *Optimization Methods and Software*, *21*(4), 527–540.

Ge, Y., Zhao, S., Zhou, H., Pei, C., Sun, F., Ou, W., & Zhang, Y. (2020). Understanding Echo Chambers in E-commerce Recommender Systems. *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2261–2270.

Ghafoori, Z., Erfani, S. M., Rajasegarar, S., Bezdek, J. C., Karunasekera, S., & Leckie, C. (2018). Efficient Unsupervised Parameter Estimation for One-Class Support

Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(10), 5057–5070.

Guerbai, Y., Chibani, Y., & Meraihi, Y. (2021). Techniques for Selecting the Optimal Parameters of One-Class Support Vector Machine Classifier for Reduced Samples. *International Journal of Applied Metaheuristic Computing*, *13*(1), 1–15.

Han, Y., Han, Z., Wu, J., Yu, Y., Gao, S., Hua, D., & Yang, A. (2020). Artificial Intelligence Recommendation System of Cancer Rehabilitation Scheme Based on IoT Technology. *IEEE Access*, *8*, 44924–44935.

Harrou, F., Bouyeddou, B., Sun, Y., & Kadri, B. (2019). Detecting cyber-attacks using a CRPS-based monitoring approach. *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018*, 618–622.

Hurley, N., Cheng, Z., & Zhang, M. (2009). Statistical attack detection. *RecSys'09 - Proceedings of the 3rd ACM Conference on Recommender Systems*, 149–156.

Ibrahim, M. H. (2021). ODBOT: Outlier detection-based oversampling technique for imbalanced datasets learning. *Neural Computing and Applications*, *33*(22), 15781–15806.

Ji, Y., Hong, W., Shangguan, Y., Wang, H., & Ma, J. (2016). Regularized singular value decomposition in news recommendation system. *ICCSE 2016 - 11th International Conference on Computer Science and Education*, *Iccse*, 621–626.

Kandimalla, S. Y., Vamsi, D. M., Bhavani, S., & Manikandan, V. M. (2022). Recent Methods and Challenges in Brain Tumor Detection using Medical Image Processing. *Recent Patents on Engineering*, *17*.

Karasmanoglou, A. (2023). ECG-Based Semi-Supervised Anomaly Detection for Early Detection and Monitoring of Epileptic Seizures. *International Journal of Environmental Research and Public Health*, *20*(6), 5000.

Kasimbeyli, R., Kamisli, Z., Nergiz, O., Dinc, G., Banu, Y., & Erdem, I. (2017). Comparison of Some Scalarization Methods in Multiobjective Optimization Comparison of Scalarization Methods. *Bulletin of the Malaysian Mathematical Sciences Society*.

Kołcz akolcz, A., Abdur Chowdhury, ieeeorg, & Alspector, J. (n.d.). *Data duplication: an imbalance problem ?*

Kulkarni, P. V., Rai, S., & Kale, R. (2020). Recommender System in eLearning: A Survey. *Proceeding of International Conference on Computational Science and Applications*, 119–126.

Kużelewska, U., & Wichowski, K. (2015). A modified clustering algorithm DBSCAN used in a collaborative filtering recommender system for music recommendation. In *Advances in Intelligent Systems and Computing* (Vol. 365, pp. 245–254). Springer Verlag.

Lam, S. K., & Riedl, J. (2004). Shilling recommender systems for fun and profit. *Thirteenth International World Wide Web Conference Proceedings, WWW2004*, 393–402.

Lazariv, T., & Schmid, W. (2018). Challenges in Monitoring Non-stationary Time Series. In *Frontiers in Statistical Quality Control 12* (pp. 257–275). Springer, Cham.

Li, C., Mo, L., Tang, H., & Yan, R. (2020). Lifelong condition monitoring based on NB-IoT for anomaly detection of machinery equipment. *Procedia Manufacturing*, *49*, 144–149.

Li, Q., & Kim, J. (2021). A deep learning-based course recommender system for sustainable development in education. *Applied Sciences (Switzerland)*, *11*(19).

Machaka, P., Bagula, A., & Nelwamondo, F. (2017). Using exponentially weighted moving average algorithm to defend against DDoS attacks. *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference, PRASA-RobMech 2016*.

Mehta, B., & Nejdl, W. (2009). Unsupervised strategies for shilling detection and robust collaborative filtering. In *User Modeling and User-Adapted Interaction* (Vol. 19, Issues 1-2 SPEC. ISS.).

Metzger, P. T., Grundy, W. M., Sykes, M. V., Stern, A., Bell, J. F., Detelich, C. E., Runyon, K., & Summers, M. (2022). Moons are planets: Scientific usefulness versus cultural teleology in the taxonomy of planetary science. *Icarus*, *374*.

Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, *106*, 36–54.

Mobasher, B., Burke, R., Bhaumik, R., & Sandvig, J. J. (2007). Attacks and remedies in collaborative recommendation. *IEEE Intelligent Systems*.

Mobasher, B., Burke, R., Bhaumik, R., & Williams, C. (2005). Effective Attack Models for Shilling Item-Based Collaborative Filtering System. *WEBKDD*.

Moustakides, G. V., Polunchenko, A. S., & Tartakovsky, A. G. (2009). Numerical comparison of CUSUM and shiryaev-roberts procedures for detecting changes in distributions. *Communications in Statistics - Theory and Methods*, *38*(16–17), 3225–3239.

Nawara, D., & Kashef, R. (2021). Deploying Different Clustering Techniques on a Collaborative-based Movie Recommender. *1. Nawara, D., Kashef, R.: Deploying Different Clustering Techniques on a Collaborative-Based Movie Recommender. 15th Annu. IEEE Int. Syst. Conf. SysCon 2021 - Proc. 4–9 (2021).15th Annual IEEE International Systems Conference, SysCon 2021 - Proceedings*, 4–9.

O'Mahony, M. P., Hurley, N. J., & Silvestre, G. C. M. (2003). Collaborative filtering – Safe and sound? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

O'Mahony, M. P., Hurley, N. J., & Silvestre, G. C. M. (2005). Recommender systems: Attack types and strategies. *Proceedings of the National Conference on Artificial Intelligence*, *1*, 334–339.

Osanaiye, O., Alfa, A. S., & Hancke, G. P. (2018). A statistical approach to detect jamming attacks in wireless sensor networks. *Sensors (Switzerland)*, *18*(6).

Ozturk, G., & Ciftci, M. T. (2014). Clustering based polyhedral conic functions algorithm

in classification. *Journal of Industrial and Management Optimization*, *11*(3), 921–932.

Page, E. S. (1954). Continuous inspection schemes. *Oxford University Press on Behalf of Biometrika Trust*, *1*(2), 100–115.

Panagiotakis, C., Papadakis, H., & Fragopoulou, P. (2020). Unsupervised and supervised methods for the detection of hurriedly created profiles in recommender systems. *International Journal of Machine Learning and Cybernetics*, *11*(9), 2165–2179.

Polunchenko, A. S., & Raghavan, V. (2018). Comparative performance analysis of the Cumulative Sum chart and the Shiryaev-Roberts procedure for detecting changes in autocorrelated data. *Applied Stochastic Models in Business and Industry*, *34*(6), 922–948.

Puglisi, S., Parra-Arnau, J., Forné, J., & Rebollo-Monedero, D. (2015). On content-based recommendation and user privacy in social-tagging systems. *Computer Standards and Interfaces*, *41*, 17–27.

Ransewa, S., Elz, N., Thanon, N., & Intajag, S. (2018). Anomaly detection using source port data with shannon entropy and EWMA control chart. *International Conference on Control, Automation and Systems*, *2018-Octob*(Iccas), 596–601.

Rezaimehr, F., & Dadkhah, C. (2020). A survey of attack detection approaches in collaborative filtering recommender systems. In *Artificial Intelligence Review* (Issue 0123456789). Springer Netherlands.

Rezaimehr, F., & Dadkhah, C. (2021). A survey of attack detection approaches in collaborative filtering recommender systems. *Artificial Intelligence Review*, *54*(3), 2011–2066.

Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, *1*(3), 239–250.

Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (1998). Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery*, *2*(2), 169–194.

Sardianos, C., Varlamis, I., Dimitrakopoulos, G., Anagnostopoulos, D., Alsalemi, A., Bensaali, F., Himeur, Y., & Amira, A. (2020). REHAB-C: Recommendations for Energy HABits Change. *Future Generation Computer Systems*, *112*, 394–407.

Serrano-Guerrero, J., Herrera-Viedma, E., Olivas, J. A., Cerezo, A., & Romero, F. P. (2011). A google wave-based fuzzy recommender system to disseminate information in University Digital Libraries 2.0. *Information Sciences*, *181*(9), 1503–1516.

Shauly-Aharonov, M., Pollak, M., & Plakht, Y. (2017). A sequential method for real-time detection of life-threatening signals in serum potassium level after myocardial infarction. *Sequential Analysis*, *36*(2), 179–193.

Shewhart, W. A. (1926). Quality Control Charts. *Bell System Technical Journal*, *5*(4), 593–603.

Shiryaev, A. N. (1961). The problem of quickest detection of a violation of stationary

behavior. *Doklady Akademii Nauk SSSR*, *138*(5), 1039–1042.

Si, M., & Li, Q. (2020a). Shilling attacks against collaborative recommender systems: a review. *Artificial Intelligence Review*, *53*(1), 291–319.

Si, M., & Li, Q. (2020b). Shilling attacks against collaborative recommender systems: a review. *Artificial Intelligence Review*, *53*(1), 291–319.

Sklavounos, D., Edoh, A., & Plytas, M. (2017). A Statistical Approach Based on EWMA and CUSUM Control Charts for R2L Intrusion Detection. *Proceedings - 2017 Cybersecurity and Cyberforensics Conference, CCC 2017*, *2018-Septe*, 25–30.

Tartakovsky, A. G., Polunchenko, A. S., & Sokolov, G. (2013). Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal on Selected Topics in Signal Processing*, *7*(1), 4–11.

Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, *50*(1), 21–48.

Torres, N., & Mendoza, M. (2019). Clustering approaches for top-k recommender systems. *International Journal on Artificial Intelligence Tools*, *28*(5), 1–27.

van Zyl, C., & Lombard, F. (2022). Signed sequential rank Shiryaev–Roberts schemes. *Quality and Reliability Engineering International*, *38*(3), 1251–1271.

Wang, S., Liu, Q., Zhu, E., Porikli, F., & Yin, J. (2018). Hyperparameter selection of one-class support vector machine by self-adaptive data shifting. *Pattern Recognition*, *74*, 198–211.

Wang, S., Liu, Q., Zhu, E., Yin, J., & Zhao, W. (2017). MST-GEN: An Efficient Parameter Selection Method for One-Class Extreme Learning Machine. *IEEE Transactions on Cybernetics*, *47*(10), 3266–3279.

Wang, Y., Zhang, R., Masoud, N., & Liu, H. X. (2023). Anomaly detection and string stability analysis in connected automated vehicular platoons. *Transportation Research Part C: Emerging Technologies*, *151*(February), 104114.

Williams, C., & Mobasher, B. (2006). *Thesis: Profile Injection Attack Detection for Securing Collaborative Recommender Systems*. *January*.

Wu, Y. (2020). Estimation of common change point and isolation of changed panels after sequential detection. *Sequential Analysis*, *39*(1), 52–64.

Wu, Y., & Wu, W. B. (2022). Sequential common change detection, isolation, and estimation in multiple poisson processes. *Sequential Analysis*, *41*(2), 176–197.

Wu, Z., Wang, Y., Wang, Y., Wu, J., Cao, J., & Zhang, L. (2016). Spammers detection from product reviews: A hybrid model. *Proceedings - IEEE International Conference on Data Mining, ICDM*.

Wu, Z., Wu, J., Cao, J., & Tao, D. (2012). HySAD: A semi-supervised hybrid shilling attack detector for trustworthy product recommendation. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Xia, H., Fang, B., Gao, M., Ma, H., Tang, Y., & Wen, J. (2015). A novel item anomaly

detection approach against shilling attacks in collaborative recommendation systems using the dynamic time interval segmentation technique. *Information Sciences*, *306*, 150–165.

Xiao, Y., Wang, H., & Xu, W. (2015). Parameter selection of gaussian kernel for one-class SVM. *IEEE Transactions on Cybernetics*, *45*(5), 941–953.

Xu, Y., & Zhang, F. (2019). Detecting shilling attacks in social recommender systems based on time series analysis and trust features. *Knowledge-Based Systems*, *178*, 25–47.

Yang, C. C., & Jiang, L. (2018). Enriching User Experience in Online Health Communities Through Thread Recommendations and Heterogeneous Information Network Mining. *IEEE Transactions on Computational Social Systems*, *5*(4), 1049–1060.

Yang, L., Huang, W., & Niu, X. (2017). Defending shilling attacks in recommender systems using soft co-clustering. *IET Information Security*, *11*(6), 319–325.

Yang, Z., Cai, Z., & Guan, X. (2016). Estimating user behavior toward detecting anomalous ratings in rating systems. *Knowledge-Based Systems*.

Yang, Z., Xu, L., Cai, Z., & Xu, Z. (2016). Re-scale AdaBoost for attack detection in collaborative filtering recommender systems. *Knowledge-Based Systems*, *100*, 74–88.

Yu, D., Mukherjee, A., Li, J., Jin, L., Wen, K., & Zhang, J. (2022). Performance of the Shiryaev-Roberts-type scheme in comparison to the CUSUM and EWMA schemes in monitoring weibull scale parameter based on Type I censored data. *Quality and Reliability Engineering International*, *February*, 3379–3403.

Yue, X., Mo, X., Wang, C., & Yao, X. (2017). Research on real-time flow abnormal traffic detection system based on DDoS attack. *Advances in Intelligent Systems and Computing*, *541*, 206–212.

Zahra, S., Ghazanfar, M. A., Khalid, A., Azam, M. A., Naeem, U., & Prugel-Bennett, A. (2015). Novel centroid selection approaches for KMeans-clustering based recommender systems. *Information Sciences*, *320*, 156–189.

Zapata, M., Madrenas, J., Zapata, M., Alvarez, J., & Chalco. (2018). Hierarchical Clustering for Collaborative Filtering Recommender Systems. *Advances in Artificial Intelligence, Software and Systems Engineering*, *787*(7), 284–292.

Zhang, F., & Wang, S. (2020). Detecting Group Shilling Attacks in Online Recommender Systems Based on Bisecting K-Means Clustering. *IEEE Transactions on Computational Social Systems*, *7*(5), 1189–1199.

Zhang, F., Zhang, Z., Zhang, P., & Wang, S. (2018). UD-HMM: An unsupervised method for shilling attack detection based on hidden Markov model and hierarchical clustering. *Knowledge-Based Systems*, *148*, 146–166.

Zhang, F., & Zhou, Q. (2014). HHT-SVM: An online method for detecting profile injection attacks in collaborative recommender systems. *Knowledge-Based Systems*.

Zhang, J., Zou, C., & Wang, Z. (2011). A new chart for detecting the process mean and

variability. *Communications in Statistics: Simulation and Computation*, *40*(5), 728–743.

Zhang, L., Yuan, Y., Wu, Z., & Cao, J. (2017). Semi-SGD: Semi-Supervised Learning Based Spammer Group Detection in Product Reviews. *Proceedings - 5th International Conference on Advanced Cloud and Big Data, CBD 2017*, 368–373.

Zhang, S., Chakrabarti, A., Ford, J., & Makedon, F. (2006). Attack detection in time series for recommender systems. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *2006*, 809–814.

Zhang, S., Ouyang, Y., Ford, J., & Makedon, F. (2006). Analysis of a low-dimensional linear model under recommendation attacks. *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *2006*(January), 517–524.

Zhang, W., & Mei, Y. (2022). Bandit Change-Point Detection for Real-Time Monitoring High-Dimensional Data Under Sampling Control. *Technometrics*, *0*(0), 1–29.

Zhang, Y., Tan, Y., Zhang, M., Liu, Y., Chua, T. S., & Ma, S. (2015). Catch the black sheep: Unified framework for shilling attack detection based on fraudulent action propagation. *IJCAI International Joint Conference on Artificial Intelligence*, *2015-Janua*(Ijcai), 2408–2414.

Zheng, Y. (2018). Identifying dominators and followers in group decision making based on the personality traits. *CEUR Workshop Proceedings*, *2068*.

Zhou, Q., & Duan, L. (2021). Semi-supervised recommendation attack detection based on Co-Forest. *Computers and Security*, *109*, 102390.

Zhou, W., Wen, J., Xiong, Q., Gao, M., & Zeng, J. (2016). SVM-TIA a shilling attack detection method based on SVM and target item analysis in recommender systems. *Neurocomputing*, *210*, 197–205.

ÖZGEÇMİŞ

ORCID NO: 0000-0001-5547-6485

Ad Soyad : Halil İbrahim AYAZ

Yabancı Dil : İngilizce

Eğitim ve Mesleki Geçmişi:

- 2009-2014, Hacettepe Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği (Lisans)
- 2014-2018, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği (Yüksek Lisans)
- 2014-şuan, Araştırma Görevlisi, Necmettin Erbakan Üniversitesi, Endüstri Mühendisliği

Yayınları ve/veya Bilimsel/Sanatsal Faaliyetleri:

Yükseköğretim Kurumları tarafından destekli bilimsel araştırma projesi:

- AYAZ Halil İbrahim (Araştırmacı), KAMIŞLI ÖZTÜRK Zehra (Yürütücü), Öneri Sistemleri için Gürbüz Tahmin Modellerinin Geliştirilmesi

Uluslararası hakemli dergilerde yayımlanan makaleler:

- AYAZ Halil İbrahim, KAMIŞLI ÖZTÜRK Zehra (2021). A Mathematical Model and a Heuristic Approach for Train Seat Scheduling to Minimize Dwell Time. Computers & Industrial Engineering, 158, Doi: 10.1016/j.cie.2021.107590 (Yayın No: 7148360)

Uluslararası hakemli dergilerde yayımlanan makaleler:

- 2023, Konferans – Tam Metin Bildiri, AYAZ Halil İbrahim, KAMIŞLI ÖZTÜRK Zehra, Polyhedral Conic Functions with Oversampling, 3rd International Conference on Engineering and Applied Natural Sciences. 740-743, Konya
- 2022, Konferans – Tam Metin Bildiri, AYAZ Halil İbrahim, KAMIŞLI ÖZTÜRK Zehra, Online Detection of Shilling Attacks on Recommender Systems with Shiryaev Roberts Procedure. 10th International Conference on Advanced Technologies, 114-117. Van
- 2022, Konferans – Tam Metin Bildiri, AYAZ Halil İbrahim, ERVURAL Bilal, A Hybrid Fuzzy Rule-Based Polyhedral Separation Approach: Medical Diagnosis Application. Intelligent and Fuzzy Systems. INFUS 2022., 504, 73-81., Doi: 10.1007/978-3-031-09173-5_10
- 2021, Konferans – Tam Metin Bildiri, AYAZ Halil İbrahim, KAMIŞLI ÖZTÜRK Zehra, Shilling Attack Detection for Recommender Systems with Quality Tools. International Cumhuriyet Artificial Intelligence Applications Conference (CAIAC) 2021, Sivas

- 2019, Konferans – Özet Bildiri, TONGUR Vahit, AYAZ Halil İbrahim (2019). Whale Optimization Algorithm for Travelling Salesman Problem. 2. Uluslararası Erciyes Bilimsel Araştırmalar Kongresi
- 2019, Konferans – Özet Bildiri, AYAZ Halil İbrahim, TONGUR Vahit (2019). Optimization of Hyperplane Paramaters with Particle Swarm Optimization. 2. Uluslararası Erciyes Bilimsel Araştırmalar Kongresi
- 2018, Konferans – Özet Bildiri, TONGUR Vahit, YENİDOĞAN Büşra, BATIBAY Ahmet Burçin, AYAZ Halil İbrahim. Automatic Measurement of Powder Particle Sizes by Image Segmantaiton Methods. International Congress on Vocational and Technical Sciences-III (UMTEB III)
- 2018, Konferans – Tam Metin Bildiri, AYAZ Halil İbrahim, TONGUR Vahit, KARAKOYUN Murat. Assessment of Survey Results on Objective Perspective. 3. Uluslararası Mesleki ve Teknik Bilimler Kongresi, 2, 1242-1248.
- 2018, Konferans – Tam Metin Bildiri, TONGUR Vahit, KARAKOYUN Murat, AYAZ Halil İbrahim. Whale Optimization Algorithm to Solve Quadratic Assignment Problems. 3. Uluslararası Mesleki ve Teknik Bilimler Kongresi, 2, 1249-1253.
- 2018, Konferans – Özet Bildiri, KARAKOYUN Murat, TONGUR Vahit, AYAZ Halil İbrahim. Grey Wolf Optimizer (GWO) algorithm for travelling salesman problem. 3. Uluslararası Mesleki ve Teknik Bilimler Kongresi
- 2018, Konferans – Özet Bildiri, AYAZ Halil İbrahim, TONGUR Vahit. Determining Number of Optimal Milk Collection Center with Multi-Objective Optimization Methods. 3. Uluslararası Mesleki ve Teknik Bilimler Kongresi, 1, 485-485.
- 2018, Konferans – Özet Bildiri, KARAKOYUN Murat, TONGUR Vahit, AYAZ Halil İbrahim. Parameter Analysis for Shuffled Frog Leaping Algorithm (SFLA). 3. Uluslararası Mesleki ve Teknik Bilimler Kongresi (Özet Bildiri/Sözlü Sunum)(Yayın No:7142583)
- 2017, Konferans – Özet Bildiri, KARAKOYUN Murat, AYAZ Halil İbrahim, TONGUR Vahit. A Comperative Study: Missing Value Estimation Methods with Classification by K-NN Algorithm. 2nd International Energy and Engineering Conference 2017
- 2017, Konferans – Özet Bildiri, KARAKOYUN Murat, TONGUR Vahit, AYAZ Halil İbrahim, TAŞKIN Ömer Faruk. Ant Colony Optimization to Solve Two-Dimensional Cutting Plane Problems. 2nd International Energy and Engineering Conference 2017, 895-895.
- 2017, Konferans – Özet Bildiri, KARAKOYUN Murat, AYAZ Halil İbrahim, TONGUR Vahit, UYAR Elif. Image Processing For Classification Of Eggs. 2nd International Energy and Engineering Conference 2017, 894-894.
- 2017, Konferans – Özet Bildiri, AYAZ Halil İbrahim, İNAN Onur, KOÇER Sabri, GÜLCÜ Şaban. Solution of the asymmetric travelling salesman problem using the migrating birds optimization algorithm. 4th INTERNATIONAL CONFERENCE ON COMPUTATIONAL AND EXPERIMENTAL SCIENCE AND ENGINEERING (ICCESEN-2017), 184-184.

- 2017, Konferans – Tam Metin Bildiri, AYAZ Halil İbrahim, TESTİK Murat Caner. Automation of FMEA for Computer Servers Using Log Data with Grey Relational Analysis. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı (UBMK 2017), Doi: 10.1109/UBMK.2017.8093480
- 2017, Konferans – Özet Bildiri, GÜLCÜ Şaban, AYAZ Halil İbrahim. Solution Of The Symmetric Travelling Salesman Problem Using The Migrating Birds Optimization Algorithm. 3rd International Conference on Engineering and Natural Science

Yazılan ulusal/uluslararası kitaplar veya kitaplardaki bölümler:

- Intelligent and Fuzzy Systems, Bölüm adı:(A Hybrid Fuzzy Rule-Based Polyhedral Separation Approach: Medical Diagnosis Application) (2022)., AYAZ Halil İbrahim, ERVURAL Bilal, Springer, Cham, Editör:Kahraman, C., Tolga, A.C., Cevik Onar, S., Cebi, S., Oztaysi, B., Sari, I.U, Basım sayısı:1, Sayfa Sayısı 8, ISBN:978-3-031-09173-5, İngilizce(Bilimsel Kitap),