**AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ**

**AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ**

*Əlyazması hüququnda*

**Rüfət Asim oğlu Şıxılı**

**Kamil İslam oğlu Qədirov**

**Rəşad Şövqi oğlu Səfərov**

**Xədicə Fazil qızı Məmmədli**

**Aysu Namaz qızı Musayeva**

# "KLASSİFİKATORLAR ANSAMBLI ƏSASINDA ŞƏBƏKƏ TRAFİKİNDƏ ANOMALİYALARIN AŞKARLANMASI" mövzusunda

## MAGİSTRLİK DİSSERTASİYASI

**İxtisasın şifri:** 060632-"İnformasiya texnologiyaları və sistemləri mühəndisliyi"

**İxtisaslaşma:** "Kibertəhlükəsizlik (SABAH)"

**Elmi rəhbər**: tex.f.d., dos. Məkrufə Şərif qızı Hacırəhimova

**Bakı – 2024**

**AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ**

**YÜKSƏK TƏHSİL İNSTİTUTU**

***MAGİSTRANTIN ANDI***

KLASSİFİKATORLAR ANSAMBLI ƏSASINDA ŞƏBƏKƏ TRAFİKİNDƏ ANOMALİYALARIN AŞKARLANMASI mövzusunda təqdim etdiyimiz magistrlik dissertasiyasını elmi əxlaq normalarına və istinad qaydalarına tam riayət etməklə və istifadə etdiyim bütün mənbələri ədəbiyyat siyahısında əks etdirməklə yazdığıma and içirik və magistrlik dissertasiyasının AzTU Kitabxana İnformasiya Mərkəzində saxlanılması, həmin mərkəz tərəfindən AzTU Rəqəmsal Repozitoriyasına daxil edilərək repozitoriyanın veb saytında yerləşdirilməsinə icazə veririk.

_____ Şıxılı Rüfət_____ _____

(Adı, Soyadı) (imza)

_____Qədirov Kamil ___ _____

(Adı, Soyadı) (imza)

_____ Səfərov Rəşad____ _____

(Adı, Soyadı) (imza)

___ Məmmədli Xədicə __ _____

(Adı, Soyadı) (imza)

____ Musayeva Aysu___ _____

(Adı, Soyadı) (imza)

# XÜLASƏ

21-ci əsrdən informasiyanın artıq kağız daşıyıcılarda deyil, informasiya sistemlərində saxlanılıb istifadə olunması kiber məkanda informasiya təhlükəsizliyinin təmin olunması problemlərini ortaya çıxarmışdır. Burada əsas məsələlərdən biri də kibertəhlükəsizliyin təmin edilməsi üçün şəbəkədə anomaliyalarının aşkarlanması prosesidir.

Şəbəkədə anomaliyaların aşkarlanması müxtəlif tətbiq sahələrində bir çox tədqiqatçıların diqqətini cəlb edən fundamental tədqiqat mövzusudur. Burada əsas məqsəd şəbəkə daxilində normal və ya məlum davranışdan kənara çıxan və təhlükəsizliyə təsir göstərdiyi güman edilən anormal hadisələrin müəyyən edilməsidir. Şəbəkə sistemindəki anomaliyalar anlayışını dərk etmək üçün nəyin normal hesab edildiyini anlamaq vacibdir.

Dissertasiya işinin əsas məqsədi şəbəkə trafikində anomal verilənləri aşkarlamaq üçün geniş yayılmış maşın təlimi alqoritmlərini tətbiq etmək və anomaliyaların aşkarlanmasında dəqiqliyin yüksəldilməsi üçün klassifikator ansamblı metodunu tətbiq etməkdir. Həyata keçirilmiş sınaqlar, şəbəkə trafiki verilənlərində anomaliyaların (DoS/DDoS hücumlar) aşkarlanmasında klassifikator ansamblının istifadəsinin ayrı-ayrılıqda götürülmüş alqoritmlərin nəticələrindən daha dəqiq nəticə nümayiş etdirdiyini göstərdi.

## SUMMARY

Since the 21st century, information is no longer stored and used in paper carriers, but in information systems, which has revealed the problems of ensuring information security in cyberspace. One of the main issues here is the process of detecting network anomalies to ensure cyber security.

Network anomaly detection is a fundamental research topic that has attracted the attention of many researchers in various application fields. The main goal here is to identify anomalous events within the network that deviate from normal or known behavior and are believed to affect security. To understand the concept of anomalies in a network system, it is important to understand what is considered normal.

The main goal of the dissertation is to apply widely used machine learning algorithms to detect anomalous data in network traffic and to apply the classifier ensemble method to improve the accuracy of anomaly detection. The conducted simulations showed that the use of a classifier ensemble in detecting anomalies (DoS/DDoS attacks) in network traffic data shows a more accurate result than the results of algorithms taken separately.

# İXTİSARLARIN SİYAHISI

| | |
|---|---|
| AI | Artificial İntelligence |
| IDS | İntrusion Detection System |
| CIA | Confidentiality İntegrity Availability |
| UEBA | User and Event Behavioral Analytics |
| DNS | Domain Name System |
| DOS | Denial of Service |
| DDOS | Distributed Denial of Service |
| WEKA | Waikato Environment for Knowledge Analysis |
| ROC | Receiver Operating Characteristic |
| NB | Naive Bayes |
| TPR | True Positive Rate |
| TNR | True Negative Rate |
| FPR | False Positive Rate |
| FNR | False Negative Rate |
| R2L | Remote to Local Attack |
| U2R | User to Root Attack |
| HIDE | Hierarchical Intrusion DEtection |
| CART | Classification and Regression Tree |

# Mündəricat

# Giriş

**Mövzunun aktuallığı:** Gündəlik həyatda informasiya texnologiyalarının istifadəsinin çoxalması nəticəsində informasiya təhlükəsizliyi zərurətə çevrilmişdir. Kompüterləşdirilmiş sistemlərin kütləvi istifadəsi viruslar, mobil təhdidlər və s. kimi kritik təhlükələrə səbəb olmuşdur. Kompüter şəbəkələrinin təkamülü kompüter təhlükəsizliyi ilə bağlı narahatlıqları, xüsusən də bugünkü şəbəkə mühitində internet təhlükəsizliyi və qabaqcıl hesablama vasitələri ilə bağlı problemləri daha da gücləndirdi. Kompüter şəbəkə texnologiyalarının yaranması çox böyük imkanlar açdı, eyni zamanda qoşulmuş sistemlərin məxfiliyinə, bütövlüyünə və əlçatanlığına xələl gətirə biləcək yeni zəifliklər və hücum vektorlarını təqdim etdi. Təhlükəsiz şəbəkənin inkişafı potensial təhlükəsizlik təhdidlərini müəyyən etmək və azaltmaq üçün sistematik və vahid yanaşma tələb edir. Son illərdə kompüter müdaxiləsi hücumlarının daha mürəkkəb hala gəldiyi və trafik məlumatlarının həcminin, sürətinin və dəyişkənliyinin əhəmiyyətli dərəcədə artdığı müşahidə edilməkdədir. 2025-ci ilə qədər bəşəriyyətin kollektiv məlumatları 175 zettabayta çatacaq. Bu dataya yayımlanan video və tanışlıq proqramlarından tutmuş səhiyyə məlumat bazalarına qədər hər şey daxildir. Bütün bu məlumatların qorunması çox vacibdir. Ənənəvi üsullar və alətlər müdaxilə hücumlarının aşkarlanmasında zəiflədiyi üçün, müdaxilənin aşkarlanması sistemlərinin əksəriyyəti indi effektivlik üçün maşın öyrənmə alətləri və alqoritmlərindən istifadəni əhatə edir. Bu baxımdan şəbəkə trafikinin izlənilməsi və analizi ilə bir çox hücumların qarşısını almaq mümkündür.

Şəbəkə trafikində anomaliyaların aşkarlanması günümüzün ən vacib kibertəhlükəsizlik məsələlərindən biridir. Anomaliyaların aşkarlanması verilmiş verilənlər toplusundan anormal məlumatları aşkar edən mühüm məlumat təhlilinin həyata keçirilməsidir. Bu sahədə Vipin Kumar, Mohiuddin Ahmed kimi tədqiqatçıların tədqiqatları diqqət çəkməkdədir. Anomaliyalar mühüm hesab olunur, çünki onlar şəbəkədəki nadir hadisələri göstərir və geniş tətbiq sahələrində kritik tədbirlərin görülməsinə təkan verə bilər; məsələn, şəbəkədə qeyri-adi trafik nümunəsi kompüterin sındırıldığını və məlumatların icazəsiz istiqamətlərə ötürüldüyünü ifadə edə bilər; kredit kartı əməliyyatlarında anormal davranış - fırıldaqçılıq fəaliyyətini göstərə bilər. Anomaliyaların aşkarlanması tibbi və ictimai səhiyyə, fırıldaqçılıq

aşkarlanması, müdaxilənin aşkarlanması kimi çoxsaylı tətbiq sahələrində geniş şəkildə tətbiq edilmişdir.

Anomaliyaların aşkarlanmasında maşın təlimi metodlarından geniş şəkildə istifadə olunur. Lakin bu metodların hər birirnin müəyyən çatışmazlıqları vardır. Buna görə də klassifikatorlar ansamblı əsasında anomaliyaların aşkarlanması daha məqsədə uyğundur. Klassifikatorlar ansamblı birdən çox metodları birləşdirərək məlumatlardakı anomaliyaları müəyyən etməkdə daha effektiv bir yanaşma təmin edir. Hər bir anomaliya aşkarlama metodunun özünəməxsus xüsusiyyətləri və limitləri olduğundan, ansamblıdakı digər metodlar bu limitləri əvəz edə bilər.

Dissertasiya işinin əsas hədəfi klassifikator ansamblları əsasında anomaliyaların aşkarlanması üçün tətbiq olunan metodları müqayisəli şəkildə analiz etməkdən ibarətdir. Həmçinin dissertasiya işi çərçivəsində anomaliyaların aşkarlanmasında nəzarətli maşın təlimi alqoritmlərinin tətbiqinin mümkünlüyü eksperimental olaraq test etməkdən və klassifikasiyanın effektivliyini artırmaq məqsədi ilə bir neçə klassifikasiya alqoritmlərindən təşkil olunmuş klassifikasiya ansamblı modelini təklif etməkdir.

**Tədqiqatın məqsəd və vəzifələri:** Dissertasiya işinin əsas məqsədi şəbəkədə anomaliyaların aşkarlanmasında istifadə edilən metodlarını tədqiq etmək və şəbəkə trafikində anomal verilənləri aşkarlamaq üçün geniş yayılmış metodları tətbiq etməkdir. Məqsədə çatmaq üçün aşağıdakı məsələlər müəyyən olunmuşdur:

- İnformasiya təhlükəsizliyi və kibertəhlükəsizlik konsepsiyasının tədqiqi;
- Anomaliyaların aşkarlanmasının aktual problemləri
- Şəbəkə trafikində anomaliyaların aşkarlanması metodlarının analizi;
- Şəbəkə trafiki verilənlərində anomaliyaları aşkarlamaq üçün klassifikator ansamblı alqoritminin tətbiqi
- Eksperimentlərin aparılması.

**Tədqiqatın obyekti və predmeti:** Tədqiqatın obyekti şəbəkədə trafiki verilənləri və onlarda anomaliyaların aşkarlanması, predmeti isə geniş yayılmış maşın təlimi metodlarından ibarət klassifikatorlar ansamblı metodunun tətbiqidir. ilə aparılmışdır. Tədqiqat üzrə eksperimentlərdə WEKA proqram platforması istifadə olunmuşdur.

**Tədqiqat metodları:** Bu dissertasiya işində qoyulan məsələnin həlli zamanı informasiya təhlükəsizliyi, proqramlaşdırma üsulları, maşın öyrənmə metodlarından istifadə olunmuşdur.

**Elmi yeniliyin elementləri və praktiki həll:** Tədqiqat işində şəbəkə trafikində anomaliyaların yüksək dəqiqliklə aşkarlanması üçün klassifikatorlar ansamblı əsasında yanaşma təklif edilmiş və realizasiya edilmişdir. Təklif edilmiş yanaşmanın tətbiqi kompüter şəbəkə verilənlərində DoS/DDos kimi kiber hücumların aşkarlanmasında şəbəkənin işinə məsul şəxslər üçün faydalı ola bilər.

**Müdafiə üçün təqdim edilən nəticələr (vəzifələr):**

1. Kibertəhlükəsizlik üzrə elmi tədqiqatların müasir vəziyyətinin analizi
2. Şəbəkə trafikində anomaliyaların aşkarlanmasında istifadə olunan maşın öyrənmə metodlarının analizi
3. Ansambl metodlarının əsasları və alqoritmlərinin analizi
4. WEKA proqram platformasında şəbəkə trafiki verilənlərində anomaliyaları aşkarlamaq üçün klassifikator ansamblı alqoritminin tətbiqi

**Əsas anlayışlar:** Bu dissertasiya işinin və ümumilikdə şəbəkə trafikində anomaliyaların aşkarlanmasının iş prinsipinin anlaşılması üçün lazım olan əsas anlayışlar aşağıdakılardır:

Anomaliya - Anomaliya qeyri-adi olan və bəzi normallıq anlayışından kənara çıxan hallar və hallar qrupuna aid edilir.

Anomaliyanın aşkarlanması - Məlumatların təhlilində anomaliya aşkarlanması (həmçinin kənar göstəricilərin aşkarlanması) ümumiyyətlə məlumatların əksəriyyətindən əhəmiyyətli dərəcədə kənara çıxan və normal müəyyən edilmiş standartlara uyğun gəlməyən nadir hadisələrin və ya müşahidələrin müəyyən edilməsi kimi başa düşülür .

Maşın öyrənmə - Maşın öyrənməsi süni intellektin və kompüter elmlərinin bir sahəsi olub, verilənlərdən öyrənib və məlumatları ümumiləşdirə bilən və açıq göstərişlər olmadan tapşırıqları yerinə yetirə bilən statistik alqoritmlərin inkişafı və öyrənilməsi ilə əlaqəli sahəsidir

Ansambl metodları - Ansambl metodları tək bir modeldən istifadə etmək əvəzinə birdən çox modeli birləşdirərək modellərdə nəticələrin dəqiqliyini artırmaq məqsədi

daşıyan texnikalardır. Birləşdirilmiş modellər nəticələrin dəqiqliyini əhəmiyyətli dərəcədə artırır. Bu, maşın öyrənməsində ansambl metodlarının populyarlığını artırdı.

Klassifikator - Klassifikator məlumatların əvvəlcədən müəyyən edilmiş kateqoriyalara salınmasını və qruplaşdırılmasını həyata keçirir.

Klassifikator ansamblı - Klassifikator ansamblı hər bir maşın öyrənmə modellərinin ayrı-ayrılıqda göstərəcəyi performansdan daha yaxşı və yüksək dəqiqlikdə nəticə almaq üçün bu modelləri birləşdirir.

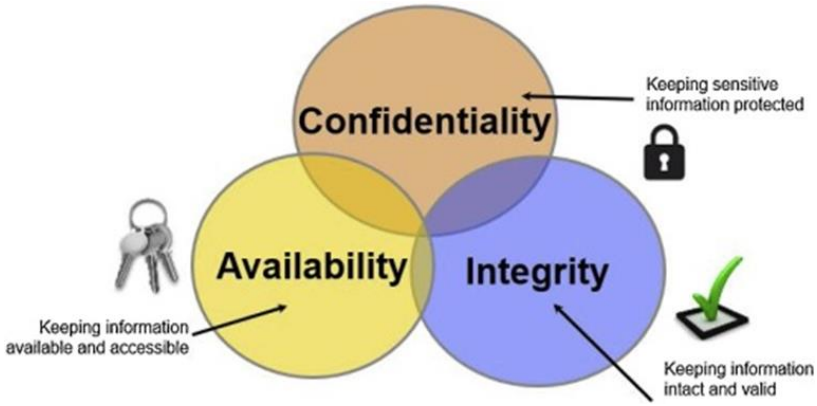
# I FƏSİL. KİBERTƏHLÜKƏSİZLİK ÜZRƏ ELMİ TƏDQİQATLARININ MÜASİR VƏZİYYƏTİNİN ANALİZİ

## 1.1 İnformasiya təhlükəsizliyi və kibertəhlükəsizlik konsepsiyası

İnformasiya təhlükəsizliyi hələ kompüter ixtirasından əvvəl yaranmışdır. Rusell və Gangemi iddia edirlər ki, informasiya təhlükəsizliyi informasiyanın özü qədər qədimdir. İnformasiya ötürülməyə, saxlanmağa və işlənməyə başladığı vaxtdan onun qorunması tələb olunurdu. Bu, insanların ilk dəfə yazmağı öyrəndiyi dövrə təsadüf edir. 1940-cı illərdən 1950-ci illərə qədər kompüterin yaranmasının başlanğıcı oldu, ilk nəsil kompüterlər yarandı. Böyük kompüterlər təcrid olunmuş müstəqil vahidlər idi və o zaman şəbəkələr mövcud deyildi. Proqramları və onların məlumatlarını kompüterlər arasında ötürmək üçün insan messencerləri və ya fiziki poçtdan istifadə edilirdi. Məlumatın ötürülməsi ilə bağlı yeganə təhlükə yaddaş daşıyıcılarının itirilə və ya oğurlana bilməsi idi. 1960-cı illərin sonu 1970-ci illərin əvvəllərində mini kompüterlərin yaradılması ilə kompüter şəbəkələrinə keçid edildi, istifadəçilərə uzaqdan verilənlərə icazəsiz daxil olmaq və istifadə etmək imkanı yarandı. İdentifikasiya və autentifikasiya da məhz həmin illərdə işə düşdü. 21-ci əsrə qədəm qoyduqda isə bir çox şey dəyişməyə başladı. Hücumçuların maliyyə qazancları üçün hakerlik etməyi kəskin şəkildə artdı. Bu, sosial mühəndislik, fişing və s. kimi təhdidlər ilə nəticələndi ki, belə halda informasiya təhlükəsizliyi məsələsi daha da aktual məsləyə çevrildi. 2020-ci ildə COVİD-19 pandemiyasından sonra sürətli rəqəmsallaşma kiber hücumların daha da artmasına səbəb olmuşdur (M.T. Dlamini, J.H.P. Eloff, M.M. Eloff, 2008).

İnformasiya təhlükəsizliyi informasiya risklərini azaltmaqla məlumatın qorunması təcrübəsidir. İnformasiya sistemlərinin və bu sistemlər tərəfindən emal edilən, saxlanılan və ötürülən məlumatlara icazəsiz daxil olma, istifadə, pozulma, dəyişdirilmə və ya məhv edilmə hallarından qorunmasını nəzərdə tutur. Bura həm rəqəmsal, həm də fiziki formada saxlanılan şəxsi məlumatların, maliyyə məlumatlarının və həssas və ya məxfi məlumatların qorunması da daxildir. Effektiv informasiya təhlükəsizliyi insanları, prosesləri və texnologiyanı cəlb edən hərtərəfli və çoxşaxəli yanaşma tələb edir. Öyrənməli olduğumuz ilk şey heç bir sistemin tamamilə təhlükəsiz olmamasıdır. İnkişaf etdirdiyimiz hər hansı bir təhlükəsizlik

tədbirindən yan keçmək üçün həmişə bir yol var. Ona görə də sistemimizin təklükəsizliyini təmin etmək üçün bütün riskləri minimuma endirməyə çalışmalıyıq. Bunun üçün möhkəm bir zəmin yaratmağa kömək edəcək təhlükəsizlik prinsipləri tətbiq edilir. Həmin təhlükəsizlik prinsiplərinə birlikdə CIA (Confidentiality İntegrity Availability-Məxfilik Tamlıq Bütövlük) triadası deyilir və bu triada üç əsas prinsipi birləşdirir: məxfilik, tamlıq və əlçatanlıq (Nina Godbole, Sunit Belpure, 2020).




**Şək. 1.1** CIA triadı (Merrill Warkentin, Craig Orgeron, 2020)

**Məxfilik** məlumatın yalnız onu görmək hüququ olan insanlar üçün əlçatan olmasına nail olmaq deməkdir və məlumatlarımızı ona baxmaq səlahiyyəti olmayan şəxslərdən qorumaq qabiliyyətimizə istinad edir. Bu onu göstərir ki, məxfi məlumatlara giriş müstəsna olaraq rəsmi səlahiyyətli şəxslər və ya qurumlarla məhdudlaşdırılmalıdır. Məxfilik həmçinin həssas məlumatların icazəsiz girişdən, açıqlanmasından və ya ifşa edilməsindən qorunmasına aiddir (Cabric, M. 2015).

**Tamlıq** məlumatın toxunulmaz və dəyişməz qalmasını təmin etmək deməkdir. Bu o deməkdir ki, məlumatlar icazəsiz şəkildə dəyişdirilə bilməz. Tamlığın pozulması çox vaxt qəsdən edilir. Təcavüzkar müdaxilə aşkarlama sistemini (IDS) keçə, icazəsiz girişə icazə vermək üçün fayl konfiqurasiyalarını dəyişdirə və ya hücumu gizlətmək üçün sistem tərəfindən saxlanılan qeydləri dəyişdirə bilər. Tamlıq bəzi hallarda təsadüfən də pozula bilər. Kimsə təsadüfən səhv kodu daxil edə və ya başqa bir ehtiyatsız səhvə yol verə bilər. Həmçinin, əgər şirkətin təhlükəsizlik siyasəti, müdafiəsi və prosedurları qeyri-adekvat olarsa, təşkilatda hər hansı bir şəxs məsuliyyət daşımadan tamlığı pozula bilər. Məlumatların tamlığını qorumaq üçün

hash funksiyalardan, şifrələmə, rəqəmsal sertifikatlar və ya rəqəmsal imzalardan istifadə edə edilir (fortinet company).

**Əlçatanlıq** o deməkdir ki, informasiya əldə etmək hüququ olan istifadəçilərin ehtiyac duyduqları zaman sistemdən məlumat əldə edə bilərlər. Yəni heç bir hal istifadəçinin qanuni və vaxtında məlumat əldə etməsinə mane olmamalıdır. Əlçatanlığın itirilməsi məlumatlarımıza daxil olmağımıza imkan verən sistemin istənilən yerində geniş çeşidli fasilələrə aid ola bilər. Bu cür problemlər enerji itkisi, əməliyyat sistemi və ya proqram təminatı problemləri, şəbəkə hücumları və ya digər problemlər nəticəsində yarana bilər (Andress, J. 2014).

İnformasiya təhlükəsizliyinin təmin edilməsində təhlükəsizlik modellərinin böyük rolu vardır. Təhlükəsizlik modelləri məlumatın qorunmasına mütəşəkkil və metodik yanaşma təklif edir, məlumatların məxfiliyinə, tamlığına və əlçatanlığına zəmanət verir ki, bu da məlumat təminatının artırılmasına kömək edir. Şəbəkə və kibertəhlükəsizlik davamlı olaraq inkişaf edən sahələr olduğundan, zamanın tarixində çoxlu təhlükəsizlik modelləri təklif edilmişdir. Bununla belə, bir çox digər modellərin əsasını təşkil edən üç klassik təhlükəsizlik modeli var:

**1.Bell-LaPadula modeli** - Kompüter təhlükəsizliyi sahəsində qabaqcıl olan David Bell və Leonard LaPadula 1970-ci illərdə Bell-LaPadula modelini yaratdılar. Bell-LaPadula yalnız öz təhlükəsizlik səviyyəsində və ya ondan yuxarı olan istifadəçilərə məzmun yaratmağa icazə verir. Bununla belə, istifadəçilər öz təhlükəsizlik səviyyələrində və ya ondan aşağı olan hər şeyi görməklə məhdudlaşır.

Bell-LaPadula modelinin qaydaları:

*Sadə məxfilik qaydası (Simple Confidentiality Rule)*: Bu qayda müəyyən edir ki, subyekt yalnız eyni məxfilik təbəqəsi və məxfiliyin aşağı təbəqəsi ilə qorunan sənədləri oxuya bilər.

*Ulduz məxfilik qaydası (Star Confidentiality Rule)*: Bu qaydaya əsasən, subyekt yalnız eyni və yuxarı məxfilik səviyyəsinə malik fayl yaza bilər.

*Güclü ulduz məxfilik qaydası (Strong Star Confidentiality Rule)*: Güclü ulduz məxfilik qaydası ən güclü və ən təhlükəsizdir, bu qayda subyektin yalnız eyni məxfilik səviyyəsində faylları oxuya və yaza bilər.

**2. Biba modeli -** Bell-LaPadula modeli bir çox təhlükəsizlik modellərinin yaradılmasına təsir göstərmişdir. Ancaq məxfiliyi təmin edən Bell-LaPadula modelində, məlumatların tamlığı paradiqması təmin edilmir. Bu çatışmazlığı aradan qaldırmaq üçün Biba modeli təklif olunmuşdur. Biba modelinin qaydaları:

*No Write-Up*: Bu qaydaya görə, istifadəçiyə daha aşağı tamlıq səviyyəsinə malik olan məlumatları əlavə etmək və ya dəyişmək icazəsi verilmir.

*No Read Down*: İstifadəçi bu qaydaya uyğun olaraq daha yüksək tamlıq səviyyəsinə malik elementi oxuya bilməz. Bu, istifadəçiyə yüksək tamlıq səviyyəsinə malik məlumatları görmək və ya oxumaq icazəsi olmadığını göstərir.

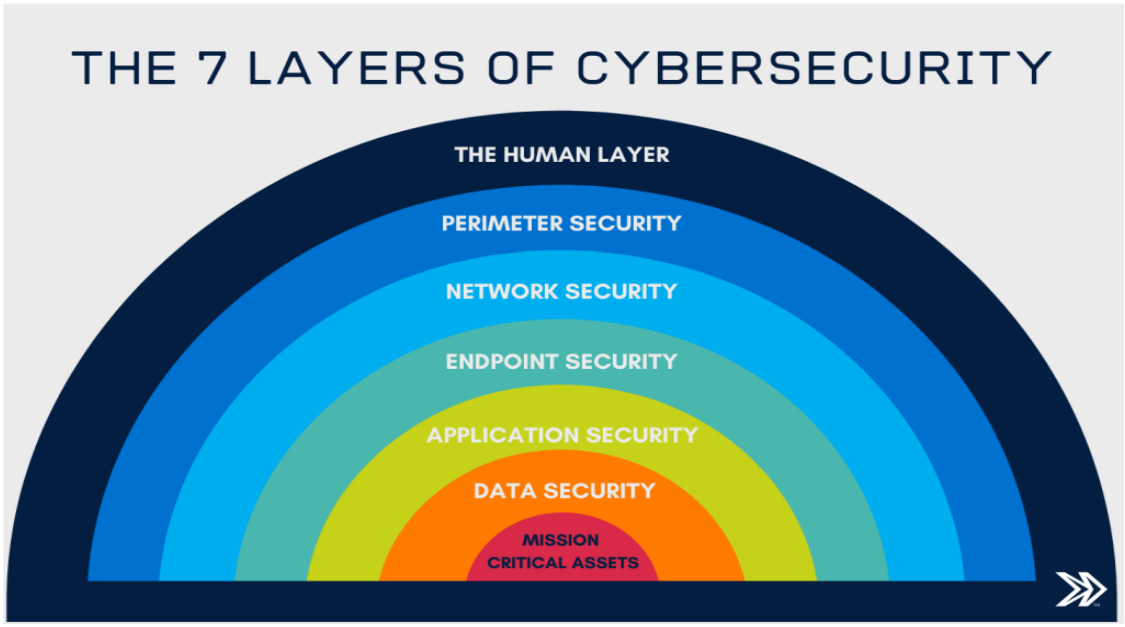
**3. Clark-Wilson modeli** - məlumatın tamlığının düşməncəsinə dəyişdirilməsi cəhdlərindən qorumaq üzərində qurulub. Bu model, sistem daxili və xarici məlumatlar arasında uyğunluğu saxlamaq və yalnız səlahiyyətli istifadəçilər məlumatları yarada və dəyişdirə bilmələrinə imkan verir.

Yuxarıda adlarını çəkdiyimiz modellərdən başqa Brewer and Nash, Harrison Ruzzo Ullman modellərini də qeyd etmək olar (Payal Wadhwa,2024).

Son dövrlərdə informasiyanın artıq kağız daşıyıcılarda deyil, informasiya sistemlərində saxlanılıb istifadə olunması kiber məkanda informasiya təhlükəsizliyinin təmin olunması problemlərini ortaya çıxarmışdır. İnformasiya təhlükəsizliyinin bir qolu olan kibertəhlükəsizlik qlobal maraq və əhəmiyyət kəsb edən məsələyə çevrilmişdir. Artıq 50-dən çox ölkə rəsmi olaraq kiberməkan, kibercinayət və kibertəhlükəsizliklə bağlı rəsmi mövqelərini əks etdirən bir növ strategiya sənədini dərc etmişdir.

Kibertəhlükəsizlik cihazlarda, şəbəkələrdə və informasiya sistemlərindəki məxfi məlumatların icazəsiz girişdən, oğurluqdan və ya zədələnmədən qorumaq təcrübəsinə aiddir. Bu, kiberhücumlardan, məlumatların pozulmasından və onlayn təhdidlərin digər formalarından qorunmaq üçün tədbirlərin həyata keçirilməsini əhatə edir.

Elmi ədəbiyyatlarda kibertəhlükəsizliyin yeddi səviyyəsi şəkil 1.2-də olduğu kimi təqdim edilmişdir .




**Şək. 1.2** Kibertəhlükəsizlik səviyyələri (Nina Godbole, SunitBelpure,2020)

1. Mission Critical Assets - qorumağınız lazım olan məlumatlardır.
2. Data Security - məlumatların saxlanmasını və ötürülməsini qoruyur.
3. Application Security - proqrama girişi, kritik aktivlərə çıxışı və daxili təhlükəsizliyi qoruyur.
4. Endpoint Security - cihazlar və şəbəkə arasındakı əlaqəni qoruyur.
5. Network Security - təşkilatın şəbəkəsini qoruyur və şəbəkəyə icazəsiz girişin qarşısını alır.
6. Perimeter Security - ümumilikdə biznesi qoruyan fiziki və rəqəmsal təhlükəsizlik metodologiyaları daxildir.
7. The Human Layer – bu səviyyəyə fişinq simulyasiyaları, zərərli insayderlər daxil olmaqla müxtəlif insan təhdidlərindən qoruyan giriş idarəetmə nəzarətləri daxildir (Nina Godbole,SunitBelpure,2020).

## 1.2 Kibertəhlükəsizlikdə süni intellekt texnologiyalarının tətbiqi imkanları

Süni intellekt (Artificial İntelligence-AI) kiber dünyada səs-küylü sözdür. 21-ci əsrin çağırışlarına uyğun olaraq, hələ də bir çox aspektlərdə inkişaf edən bir elmdir. AI-ın istifadəsi insan həyatında ayrılmaz hala gəlmişdir. Bu gün dünyanı AI olmadan təsəvvür etmək mümkün deyil, çünki bu, insan həyatına çox əhəmiyyətli dərəcədə təsir göstərir. AI-ın əsas məqsədi problemləri həll etmək üçün insan biliyini təmsil

edən texnologiyaya əsaslanan fəaliyyətləri inkişaf etdirməkdir. İnsan həyatı, oyun, nitqin tanınması, ekspert sistemi, görmə sistemi, əl yazısının tanınması, kəşfiyyat robotları, maliyyə əməliyyatları və s. AI-ın bir alt hissəsinə çevrilmişdir (Arab Mohammed Shamiulla, 2019).

AI müxtəlif sahələrdə bir çox üstünlüklər və tətbiqlər təqdim edir. Kibertəhlükəsizlik də həmin sahələrdən biridir. Sürətlə inkişaf edən kiberhücumlar və bu gün baş verən cihazların sürətlə çoxalması ilə AI kibercinayətkarlarla ayaqlaşmağa, təhlükənin aşkarlanmasını avtomatlaşdırmağa və ənənəvi proqram təminatı və ya əl(manual) üsullarından daha effektiv cavab verməyə kömək edə bilər. Kibertəhlükəsizlikdə AI-dən istifadənin bir neçə üstünlükləri və tətbiqləri bunlardır:


**Şək. 1.3** Kibertəhlükəsizlikdə süni intellektin tətbiq dairələri (Aysu Musayeva,2024)

**Threat hunting (Təhlükə ovçuluğu)** - Keçmişdə kibertəhlükəsizlik geniş şəkildə təhdidləri aşkar etmək üçün imza əsaslı üsullara etibar edirdi. Bu üsullar kiberhücumların xüsusi kateqoriyaları ilə əlaqəli təkrarlanan nümunələri və ya "imzaları" müəyyən etməklə məlum təhlükələrlə effektiv şəkildə mübarizə aparırdı. Bu yanaşma uyğun gəlməyən yeni, naməlum təhlükələri aşkar etmək və onlara dərhal cavab vermək iqtidarına malik olmayan məhdudiyyətə malikdir. Süni intellektin istifadəsinin faydası buradadır. AI, təhdidlərin aşkarlanması və identifikasiyasını təkmilləşdirmək üçün güclü proqnozlaşdırma imkanlarından istifadə edərək təhlükə ovçuluğunda inqilab edir. AI-ın gücü böyük miqdarda məlumatı səmərəli şəkildə emal etmək və təhlil etmək, mənalı nümunələri tanımaq və uyğun olmayanları aradan qaldırmaq qabiliyyətindədir. Real vaxt rejimində anomaliyaların aşkarlanması üçün etalon rolunu oynayır, burada normadan hər hansı bir kənaraçıxma potensial

təhlükəsizlik riskini göstərə bilər. Davranış analizinin AI-a daxil edilməsi təhlükənin aşkarlanmasına daha dinamik yanaşma imkanı verir. Bu metoddan istifadə edərək AI sistemləri təşkilatın şəbəkəsi daxilində çoxsaylı son nöqtə məlumatlarını emal edə bilər.

**Vulnerability Management (Zəifliyin idarə edilməsi)** -Təşkilatlar getdikcə artan sayda potensial zəifliklərlə qarşılaşır və onları effektiv şəkildə idarə etmək üçün tez-tez mübarizə aparırlar. Zəifliyin idarə olunmasına adətən reaktiv paradiqmaya əməl edən və tez-tez yüksək riskli zəifliklərin onlara müraciət etməzdən əvvəl istismar edilməsini gözləyən ənənəvi yanaşmaların mövcud kibertəhlükəsizlik mühitində qeyri-kafi olduğu sübut edilmişdir. Bu kontekstdə zəifliklərin idarə edilməsində Süni İntellektin funksiyası transformativ olur. İstifadəçi və hadisə davranış analitikası (User and Event Behavioral Analytics - UEBA) əhəmiyyətli bir yanaşmadır. UEBA süni intellekt sistemlərinə təşkilatın istifadəçi hesablarının, son nöqtələrinin və serverlərinin əsas fəaliyyətini daim təhlil etmək və öyrənmək imkanı verir. Bu təhlil müəyyən edilmiş normadan kənara çıxan davranışları müəyyən etməyə kömək edir. Bu cür kənaraçıxmalar və ya anomaliyalar sıfır gün hücumlarının mövcudluğunu göstərə bilər. Aİ bu hücumları daha əvvəl müəyyən edə bilər.

**Network Security (Şəbəkə Təhlükəsizliyi)** - Şəbəkə təhlükəsizliyi istənilən kibertəhlükəsizlik strategiyasının kritik aspekti olaraq qalır. Təhlükəsizlik siyasətləri şəbəkə təhlükəsizliyində əvəzolunmaz rol oynayır, hansı şəbəkə əlaqələrinin qanuni olduğunu və hansı potensial zərərli fəaliyyətə görə əlavə araşdırma tələb etdiyini müəyyən etməyə kömək edir. Aİ bu siyasətlərin formalaşdırılmasını əhəmiyyətli dərəcədə təkmilləşdirə bilər. Aİ-ın böyük həcmdə məlumatları təhlil etmək və nümunələrdən öyrənmək bacarığı ilə o, görünməmiş dəqiqlik və səmərəliliklə təhlükəsizlik siyasətlərinin yaradılmasını həyata keçirə bilər. Bu, daha möhkəm təhlükəsizlik vəziyyətinə, daha yaxşı təhdid identifikasiyasına və kibertəhlükələrə qarşı gücləndirilmiş qorunmaya gətirib çıxarır. Müzakirə olunan tətbiqlər Aİ-ın ənənəvi kibertəhlükəsizlik strategiyalarını necə tamamladığını və tədricən onların ayrılmaz hissəsinə çevrildiyini vurğulayır (Sarvesh Kumar, Upasana Gupta, Arvind Kumar Singh & Avadh Kishore Singh, 2023).

### 1.3 Şəbəkə hücumları və Şəbəkə hücumlarının aşkarlanması sistemləri

Bu gün bir çox məxfi və həssas məlumatlar informasiya texnologiyalarında saxlanılır və şəbəkələr vasitəsilə əldə edilə bilər. Şirkətlərə məlumatlarını qorumağa və yalnız səlahiyyətli işçilərə və qurumlara giriş icazəsi verməyə imkan verən güclü şəbəkə təhlükəsizliyinə sahib olmaq vacibdir. Məlumat təhdidlərini aşkar etmək və ya qarşısını almaq ümumiyyətlə asan deyil və bu, şirkətlərin məlumatlarını təhlükəyə atmasına və maliyyə itkisinə səbəb ola bilər. Şəbəkə təhlükəsizliyi sahəsində peşəkarlar buna görə də bu təhlükələrin qarşısını almaq və şərh etmək üçün məsuliyyət daşıyırlar. Şəbəkə hücumu uzaq kompüterin əməliyyat sisteminə daxil olmaq cəhdidir. Cinayətkarlar əməliyyat sistemi üzərində nəzarət yaratmaq, əməliyyat sisteminin xidmətdən imtinasına səbəb olmaq və ya həssas məlumatlara daxil olmaq üçün şəbəkə hücumlarına cəhd edirlər.

Cybersecurity Ventures şirkətinin verdiyi məlumata görə kibercinayətkarlığın qlobal illik dəyərinin 2024-cü ildə 9,5 trilyon ABŞ dollarına çatacağı proqnozlaşdırılmışdır. Bu, 2025-ci ilə qədər 10,5 trilyon dollara çatacağı gözlənilən kibercinayətkarlıq nəticəsində dəymiş ziyanın artmasıdır. (şək.1.4)


**Şək. 1.4** 2024 Kibercinayətkarlıq statistikası (Cybercrime Magazine)

Aparılan tədqiqatlar göstərir ki, şəbəkə hücumlarını cədvəl 1.1-də göstərilən kateqoriyalara bölmək olar:

**Cədvəl 1.1** Şəbəkə hücumlarının növləri (Aysu Musayeva, Xədicə Məmmədli,2024)

| Əsas kateqoriya | Tərif | Nümunələr |
|---|---|---|
| **Infection (İnfeksiya)** | Sistemdə zərərli fayllar quraşdırmaqla hədəf sistemi yoluxdurmağı hədəfləyir. | Virus, Worm, Trojan, Ransomware. |
| **Exploding (Daşma)** | Proqram təminatının buferə məlumat yazması buferin tutumunu aşdığı zaman baş verən şəbəkə hücumudur. | Buffer Overflow. |
| **Cheat (Aldatma)** | Bu kateqoriyanın tipik nümunələrinə saxta şəxsiyyətdən istifadə cəhdləri daxildir. | IP Spoofing, MAC Spoofing, DNS Spoofing, Session Hijacking, XSS Attacks. |
| **Password attacks (parol hücumları)** | Təcavüzkarın parolla qorunan sistemi sındırmaq üçün istifadə etdiyi hücum vektorudur. | Brute Force, Dictionary Attacks. |
| **Denial of Service (Xidmətdən imtina hücumu)** | Sistemin təmin edə biləcəyi imkanları aşan çoxlu eyni sorğular göndərməklə sistemi qurban verir. | Flooding, DDoS (Distributed Denial of Service). |

**Worm (Soxulcanlar)**: Wormlar müstəqil, yəni başqa proqramlara yeridilmədən öz surətlərini kompüter sistemlərində yaymağa və onları işə salmağa qabil olan proqramlardır (İmamverdiyev Y.N, 2015).

**DOS/DDOS hücumları (Xidmətdən imtina hücumu):** Xidmətdən imtina (DoS) və paylanmış xidmətdən imtina (DDoS) hücumlarında təcavüzkar hədəfə çoxlu sayda böyük ölçülü məlumat sorğuları göndərir. O qədər sorğular göndərilir ki, hədəf sistem həddən artıq yüklənir və xidmət üçün qanuni sorğulara cavab verə bilmir. Bunun nəticəsində sistem çökə bilər və ya sadəcə adi funksiyalarını belə yerinə yetirə bilməyəcək vəziyyətə gələ bilər. DOS və DDOS hücumları arasındakı fərq odur ki, DDOS hücumunda eyni anda bir çox cihazdan bir hədəfə qarşı koordinasiyalı sorğu

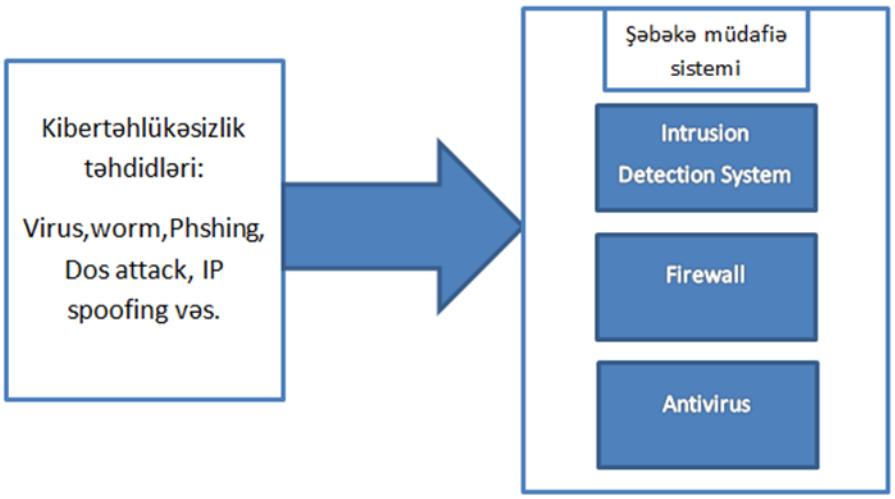
axını başlayır. Hər iki hücum növü də sistemin normal işini pozan, istifadəçilərin sistemlə əlaqə qurmasını və ya resurslardan istifadə etməsinin qarşısını alır (Michael E. Whitman & Herbert J. Mattord, 2012).

**IP Spoofing (IP saxtakarlığı):** IP saxtakarlığı etibarlı IP ünvanını təqlid edərək sistemə icazəsiz giriş əldə etmək üçün istifadə edilən bir hücumdur. Bu hücumda təcavüzkar kompüterə gələn məlumatın etibarlı sistemdən gəldiyini göstərən mesajlar göndərir (Sharmin Rashid, Subhra Prosun Paul, 2013).

**Fişinq Hücumları:** Zərərli bir aktyor etibarlı, qanuni mənbələrdən gələn kimi görünən e-məktublar göndərdikdə fişinq hücumu baş verir. Fişinq hücumları sosial mühəndisliyi və texnologiyanı birləşdirir. Hücumu həyata keçirmək üçün zərərli aktyor istifadəçini veb-sayta yönləndirən bir keçid göndərə bilər ki, bu da istifadəçini viruslar kimi zərərli proqramları yükləməyə və ya istifadəçini şəxsi məlumatlarını vermək üçün aldadır (Surbhi Gupta, Abhishek Singhal & Akanksha Kapoor, 2016).

**DNS Spoofing hücumu (Domen adı aldatma hücumu):** DNS Spoofing hücumu ilə haker saxta vebsayta trafiki yönləndirmək üçün DNS qeydlərini dəyişir. Saxta sayta daxil olduqdan sonra qurban haker tərəfindən istifadə edilə həssas məlumatları daxil edə bilər. DNS saxtakarlığı hücumunda təcavüzkar istifadəçinin daxil olduğu saytın qanuni olduğunu düşünməsindən istifadə edir (Maksutov, A. A., Cherepanov, I. A., & Alekseev, M. S. 2017).

Kiberinfrastrukturun qəsdən və potensial zərərli təhdidlərə qarşı təhlükəsizliyini təmin etmək üçün kibertəhlükəsizlik üzrə peşəkarlar və institutlar, özəl sənayelər, akademiyalar və dövlət qurumlarından olan tədqiqatçılar arasında artan əməkdaşlıq səyləri müxtəlif şəbəkə hücumlarının aşkarlanması sistemlərinin istismarı və layihələndirilməsi ilə məşğuldur. Kibertəhlükəsizlik tədqiqatçıları və dizaynerləri kompüterləri və şəbəkələri sistemə müdaxilə etmək və ya maliyyə, tibbi və ya digər məlumatları oğurlamaq istəyən hakerlərdən qoruyan müxtəlif kibermüdafiə sistemləri vasitəsilə məlumat və məlumat idarəetmə sistemlərinin məxfiliyini, bütövlüyünü və əlçatanlığını qorumağı hədəfləyirlər. Hücumun aşkarlanması həm sistem daxilindəkilər, həm də xarici müdaxiləçilər tərəfindən kompüter sistemlərindən icazəsiz istifadənin müəyyən edilməsi problemidir və böyük məlumat dəstlərində zərərli nümunələrin aşkarlanması prosesidir.




**Şək. 1.5.** Ənənəvi kibertəhlükəsizlik sistemi (Sumeet Dua & Xian Du, 2011).

Şəkil 5-də göstərildiyi kimi, ənənəvi kibertəhlükəsizlik sistemləri dos hücumları, phising, ip spoofing, viruslar və wormlar daxil olmaqla müxtəlif kibertəhlükəsizlik təhdidlərini aradan qaldırılır. Şəbəkə müdafiə sistemlərinə müdaxilənin aşkarlanması sistemləri (İntrusion Detection System - IDS), firewalllar və antivirus proqram təminatı daxildir (Sumeet Dua & Xian Du, 2011).

**IDS** şəbəkədə zərərli fəaliyyətə nəzarət edən cihaz və ya proqramdır. IDS müdaxiləni aşkar etmək üçün şəbəkə trafikindəki məlumat paketlərini izləməklə məlumatları qoruyan şəbəkə təhlükəsizliyinin bir komponentidir. Zərərli fəaliyyətləri yoxlamaq üçün şəbəkəyə nəzarət edir və təhlükəsizlik meyarlarına cavab verməyən hadisələri şəbəkə administratoruna bildirir.

Hazırda müdaxilənin aşkarlanması üçün iki əsas yanaşma istifadə olunur: sui-istifadənin aşkarlanması və anomaliyaların aşkarlanması. Sui-istifadənin aşkarlanması həmçinin imza əsaslı və ya biliyə əsaslanan sistemlər kimi tanınır. Onlar əksər antivirus proqramları ilə eyni prinsipə əməl edirlər və müdaxilə cəhdlərini aşkar etmək üçün əvvəlki hücumlar və zəifliklər haqqında toplanmış biliklərə güvənirlər. Sui-istifadənin aşkarlanması sistemləri hostun və ya monitorinq edilən şəbəkənin cari fəaliyyətini məlum hücumların "imzaları" ilə müqayisə edir. Cari fəaliyyətlər məlum imzalardan hər hansı birinə uyğun gəlirsə, həyəcan siqnalı işə salınır.

Anomaliyanın aşkarlanması isə davranışa əsaslanan sistemlər kimi də tanınır. Onlar nəzarət edilən sistemin gözlənilən davranışlarından kənaraçıxmaları müşahidə

etməklə müdaxilələrin aşkar oluna biləcəyinə arxalanırlar. Bu "normal" davranışlar ya keçmişdə edilən bəzi müşahidələrə, ya da müxtəlif üsullarla edilən bəzi proqnozlara uyğun ola bilər. Bu "normal" nümunəyə uyğun gəlməyən hər şey anormal olaraq işarələnəcək. Buna görə də, anomaliyaların aşkarlanmasının əsas prosesi nəyin anormal olduğunu öyrənmək deyil, nəyin normal və ya gözlənilən olduğunu öyrənməkdir (Shaik Akbar, Dr.K. Nageswara Rao and Dr.J.A. 2010).

**Firewall** təşkilatın daxili şəbəkəsini digər şəbəkələrdən təcrid edən aparat və proqram təminatının birləşməsidir. Şəbəkə üzərindən keçən bəzi paketlərin keçməsinə icazə verir, bəzilərini isə bloklayır. O, qoruduğu şəbəkə sahələrindəki cihazlarda icazəsiz və ya qeyri-qanuni seansların qarşısını almaq üçün fəaliyyət göstərir. Firewalllar xarici dünyadan təsdiqlənməmiş interaktiv girişlərdən qorunmaq üçün konfiqurasiya edilmişdir. Firewall bir cüt mexanizm kimi düşünülə bilər: biri trafikin qarşısını almaq üçün, digəri isə trafikə icazə vermək üçün mövcuddur. Firewallları idarə edən administratorlar firewall qaydalarını təyin edərkən diqqətli olmalıdırlar (Firkhan Ali Bin Hamid Ali, 2011).

Şəbəkə hücumlarının aşkarlanması sistemlərindən biri də **antivirus** proqramlarıdır. Antivirus proqramlarının reallaşdırılması zamanı iki üsuldan istifadə olunur:

- antivirus bazarlarında müəyyən edilmiş kriterilərə uyğun olaraq faylların yoxlanması;

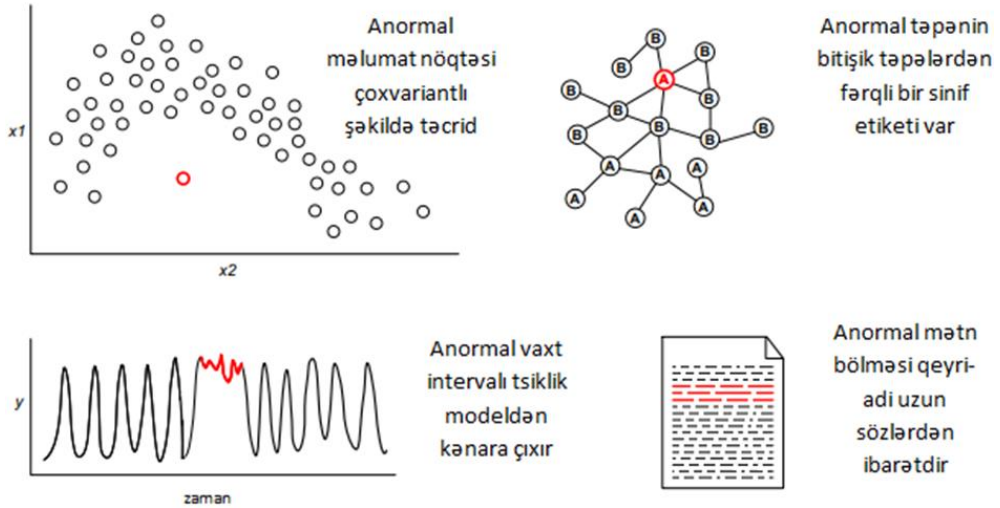
- əvvəlcədən bəlli olan yoluxmuş proqramların davranışına oxşar şübhəli davranışlı proqramların aşkar edilməsi.

Antivirus proqramı yoxlanılan faylı təhlil edərkən antivirus bazalarına müraciət edir və həmin faylın hər hansı hissəsinin kodunun bazada olan virusun koduna uyğunluğunu yoxlayır, uyğunluq aşkar edilərsə antivirus proqramı yoluxmuş faylı pozur, girişi bloklayır və ya virusun yayılmasının qarşısını almaq məqsədilə yoluxmuş faylın yerinə yetirilməsinə icazə vermir. Şübhəli davranışın aşkar edilməsi üsulu isə antivirus bazalarında olmayan yeni viruslardan qorunmanı təmin edir. Bu üsul əsasında qurulmuş bəzi proqramlar həddən artıq xəbərdarlıqlar verməklə istifadəçiləri çaşdıra bilərlər (Qasımov V.Ə. 2011).

## 1.4. Anomaliya və onun tipləri

Mövcud dövrdə kütləvi məlumatların toplanması və bunun üçün istifadə edilən təkmilləşməmiş sistemlərdə anormal müşahidələri məlumat dəstlərimizdə geniş şəkildə müşahidə edə bilərik. Anomaliyalar qeyri-adi olan və bəzi normallıq anlayışından kənara çıxan hallar və hallar qrupuna aid edilir. Bu mövzuda təxminən 250 illik nəşrlərə baxmayaraq, indiyədək müxtəlif anomaliya növlərinə dair hərtərəfli və konkret icmallar dərc olunmayıb. Anomaliyaların ümumi təriflərinin çox vaxt "qeyri-müəyyən" olduğu və tətbiq sahəsindən asılı olduğu deyilir ki, bu da çox güman ki, anomaliyaların özünü büruzə verməsinin müxtəlif yollarından asılıdır. Müxtəlif ədəbiyyatlarda anomaliya termini "kənar göstərici" (outlier), "yenilik" (novelty), "səs-küy" (noise) kimi fərqli şəkillərdə istifadə olunur (Ralph Foorthuis 2021)( Ramiz Alıguliyev, Makrufa Sh. Hajirahimova, 2019).

Anomaliyalar məlumatların təhlilinə mane olan səs- küy amili yarada bilsə də, onlar axtarılan faktiki siqnalları da təşkil edə bilər. Şəkil 1.6-da göstərildiyi kimi, onların daxil olduğu çoxlu forma və ölçülərə görə onları müəyyən etmək çətin məsələ ola bilər.



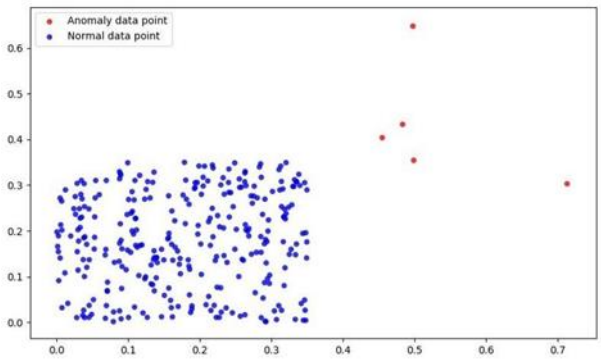

**Şək. 1.6** Anomaliyaların fərqli növləri (Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu, 2015)

Qırmızı göstərilən hadisələr anomaliyaların müxtəlifliyini göstərir və nəticədə anomaliyalar qeyri-müəyyən anlayış kimi göstərilir (Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu, 2015).

Şəbəkədə anomaliya dedikdə isə şəbəkə daxilində normal və ya məlum davranışdan kənara çıxan və təhlükəsizliyə təsir göstərdiyi güman edilən anormal

hadisələr başa düşülür, həmçinin şəbəkənin müntəzəm əməliyyatlarını pozmağa yönəlmiş hərəkətlər kimi də tanınır. Anomaliyalar verilənlərdə dəqiq müəyyən edilmiş məntiqi vəziyyət anlayışına uyğun gəlməyən nümunələrlə müəyyən edilir. Bu, sistemin hərəkətlərinin əvvəlki normal davranışından əhəmiyyətli dərəcədə kənara çıxdığı ani mərhələdir. Şəbəkə sistemindəki anomaliyalar anlayışını dərk etmək üçün nəyin normal hesab edildiyini anlamaq vacibdir. Şəbəkədə anomaliyalar iki əsas səbəbə görə baş verə bilər: performansla əlaqəli və təhlükəsizliklə əlaqəli. Şəbəkə cihazının nasazlığı, məsələn, marşrutlaşdırıcının yanlış konfiqurasiyası kimi nasazlıqlar səbəbindən performansla bağlı anomaliya baş verə bilər. Təhlükəsizliklə bağlı anomaliyalar şəbəkənin normal fəaliyyətini pozmağa cəhd edən zərərli fəaliyyətlər səbəbindən baş verir. Şəbəkə anomaliyalarının üç əsas növü var: point (nöqtə) anomaliyaları, contextual (kontekstual) anomaliyalar və collective (kollektiv) anomaliyalar.

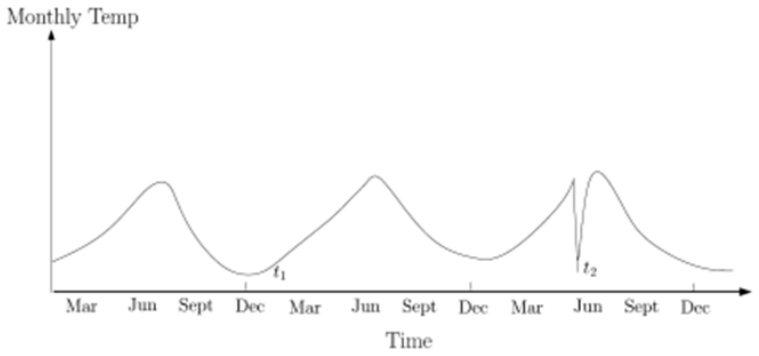
**Point anomaliya:** Müəyyən bir məlumat nümunəsi verilənlər dəstinin normal modelindən kənara çıxdıqda, bu, nöqtə anomaliyası hesab edilə bilər. Məsələn, kredit kartından istifadə etməklə adi gündəlik xərclər adətən təxminən yüz dollar təşkil edirsə, lakin müəyyən bir gündə bu rəqəm dörd yüz dollara yüksəlirsə, bu əməliyyat anormal hesab olunacaq.




**Şək. 1.7** Nöqtə anomaliyası (Davis David,2019)

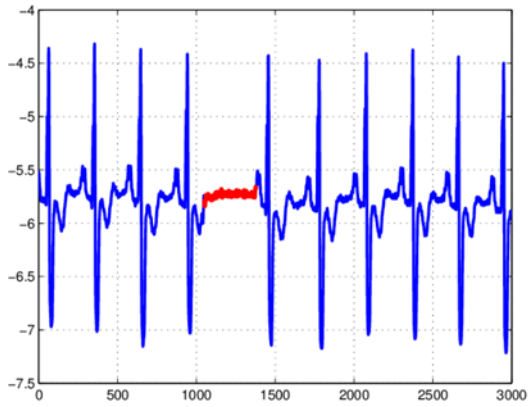
**Konstekstual anomaliya:** Məlumat müəyyən bir vəziyyətdə qeyri-adi hərəkət etdikdə buna kontekstual və ya şərti anomaliya deyilir. Bu cür anomaliya adətən zamanla dəyişən məlumatlarla əlaqələndirilir. Şəkil 1.8-də son bir neçə il ərzində ərazinin aylıq temperaturunu göstərən temperatur zaman seriyası üçün belə bir nümunə verilmişdir. Temperaturun 1°C olması qışda ($t_1$ vaxtı) normal ola bilər, lakin

yayda ($t_2$ vaxtı) eyni dəyər anomaliya hesab olunur (Varun Chandola, Arindam Banerjee & Vipin Kumar, 2009).




**Şək. 1.8** Kontekstual anomaliya (Varun Chandola, Arindam Banerjee & Vipin Kumar, 2009)

**Kollektiv anomaliya:** Kollektiv anomaliya, məlumat klasterinin məlumat dəstinin qalan hissəsi ilə müqayisədə anormal nümunələri göstərdiyi zaman baş verir. Məsələn, bir insanın elektrokardioqrammasında uzun müddət aşağı dəyərlərin olması anormal xarici bir fenomeni göstərir, halbuki özlüyündə aşağı dəyər anomal hesab edilmir (Mohammed Hussein Thwaini, 2022).




**Şək. 1.9** Kollektiv anomaliya (Mohammed Hussein Thwaini, 2022)

Nəticə olaraq, şəbəkə anomaliyaları standart davranışdan kənara çıxan anormal şəbəkə fəaliyyətləridir və onları nöqtə anomaliyalarına, kontekst anomaliyalarına və kollektiv anomaliyalara təsnif etmək olar. Bu növ anomaliyaları başa düşmək şəbəkə sistemində təhlükəsizlik təhdidlərini effektiv şəkildə aşkar etmək və azaltmaq üçün çox vacibdir.

## 1.5 Anomaliyaların aşkarlanmasının aktual problemləri

Anomaliyaların aşkarlanması müxtəlif tətbiq sahələrində bir çox tədqiqatçıların diqqətini cəlb edən fundamental tədqiqat mövzusudur. Kritik sənaye sistemlərindən - şəbəkəyə müdaxilənin aşkarlanması sistemlərindən, insanların gündəlik fəaliyyətlərinə - mobil fırıldaqçılığın aşkarlanmasına qədər, anomaliyaların aşkarlanması ictimai və şəxsi mülkləri qorumaq üçün ən vacib və ilk vasitəyə çevrilmişdir. İllər ərzində ardıcıl inkişafla, məlumatların toplanması, təmizlənməsi və inteqrasiyasının tədricən təkmilləşdirilməsi müxtəlif sahələrdə anomaliyaların aşkarlanmasını dəstəklədi. Buna baxmayaraq, məlumat həcminin kəskin artması və məlumat nümunələrinin davamlı dəyişməsi anomaliyaların aşkarlanması sistemləri üçün böyük problemlər yaradır və anomaliyaların aşkarlanmasının müxtəlif ehtiyaclarının öhdəsindən gəlmək üçün fərqli xüsusiyyətlərə malik daha ağıllı anomaliya aşkarlama metodlarının tətbiqinə tələbi artırır (Guansong Pang, Chunhua Shen, Longbing Cao & Anton Van Den Hengel, 2021).

Bir çox anomaliya aşkarlama metodları nəzəri cəhətdən cəlbedici və möhkəm olsa da, müxtəlif sahələrdə praktiki olaraq tətbiq edilməzdən əvvəl bir sıra texnoloji problemlərin aradan qaldırılması lazımdır. Bu problemlər ümumiyyətlə dəqiqliyə, səmərəliliyə və digər imkanlara, məsələn, şərh oluna bilənliyə, miqyaslılığa və s. aiddir. Araşdırmalar göstərir ki, anomaliyaların aşkarlanmasında bəzi sahələrdə problemlər mövcuddur:

***Kibertəhlükəsizlikdə anomaliyaların aşkarlanması problemləri:*** Kibertəhlükəsizlikdə anomaliyaların aşkarlanması kompüter şəbəkələri və ya sistemləri daxilində şübhəli fəaliyyətləri və ya potensial təhlükələri müəyyən etmək üçün vacibdir. Lakin anomaliyaların aşkarlanmasında bu sahədə bəzi problemlərlə üzləşirik. Bu problemlər kibertəhlükələrin dinamik xarakterindən, müasir şəbəkə infrastrukturlarının mürəkkəbliyndən və mövcud aşkarlama metodlarının məhdudiyyətlərindən irəli gəlir. Kibertəhlükəsizlikdə anomaliyaların aşkarlanması zamanı ən geniş problemlərdən olan yanlış xəbərdarlıq siqnalları həqiqi kənaraçıxmaları aşkarlamaqda çətinliklər yaradır. Bu zaman normal hallar səhv olaraq anomaliyalar kimi təqdim edilə bilər və nəticədə həqiqi anomaliyalar gözdən qaça bilər. İllər ərzində çoxlu anomaliya aşkarlama metodlarının tətbiqinə

baxmayaraq, mövcud ən müasir metodlarda hələ də yanlış xəbərdarlıq siqnalları problemi aktualdır (Varun Chandola, Arindam Banerjee &Vipin Kumar, 2009).

***Böyük məlumatlarda anomaliyaların aşkarlanması problemləri:*** Bir çox məlumat dəsti vebloglardan, maliyyə əməliyyatlarından, sağlamlıq qeydlərindən və nəzarət qeydlərindən, həmçinin biznes, telekommunikasiya və bioelmlərdən ibarətdir. Məlumat dəstlərinin böyük və paylanmış xarakterini təsvir edən termin "böyük məlumat"(big data) olaraq adlandırılır. Son illərdə böyük verilənlərin əsas problemlərindən biri də anomaliyaların aşkarlanmasıdır.

Anomaliyaların aşkarlanması anomaliyalar və ya yenilik adlanan məlumatların qalan hissəsindən kənara çıxan anormal nümunələri aşkar etmək məqsədi daşıyır. Yüksək ölçülülük anomaliyaların aşkarlanmasında çətinliklər yaradır, çünki atributların və ya xüsusiyyətlərin sayı artdıqda dəqiq ümumiləşdirmək üçün lazım olan məlumatların miqdarı da artır və nəticədə məlumat nöqtələri daha dağınıq və təcrid olunmuş məlumatların seyrəkliyi ilə nəticələnir. Məlumatların bu seyrəkliyi lazımsız dəyişənlər və ya əsl anomaliyaları gizlədən çoxsaylı uyğunsuz atributların yüksək küy səviyyəsi ilə bağlıdır. Bir çox anomaliya aşkarlama üsulları məlumat dəstlərinin yalnız bir neçə xüsusiyyətə malik olduğunu güman edir və əksəriyyəti aşağı ölçülü məlumatlarda anomaliyaları müəyyən etməyə yönəlib. Məlumat ölçüsü artdıqda yüksək ölçülülük problemini həll edən texnikalar bir sıra amillərə görə anomaliyaların aşkarlanmasında çətinliklərlə üzləşirlər (Srikanth Thudumu, Philip Branch, Jiong Jin & Jugdutt (Jack), 2020).

***Tibbi verilənlərdə anomaliyalarının aşkarlanması problemləri:*** Tibbidə anomaliyaların aşkarlanması adətən xəstəyə aid qeydlərlə (xəstənin yaşı, qan qrupu və çəkisi və s) bağlıdır. Məlumatlarda xəstənin anormal vəziyyəti, cihaz səhvləri və ya qeyd xətaları kimi bir neçə səbəbə görə anomaliyalar ola bilər. Ona görə də anomaliyaların aşkarlanması bu sahədə çox kritik problemdir və yüksək dəqiqlik tələb edir. Bu sahədə mövcud anomaliyaların aşkarlanması üsullarının əksəriyyəti anomal qeydləri (nöqtə anomaliyaları) aşkar etmək məqsədi daşıyır. Tipik olaraq etiketli məlumatlar sağlam xəstələrə aiddir, buna görə də texnikaların əksəriyyəti yarı nəzarət edilən yanaşmanı qəbul edir. Bu sahədə anomaliya aşkarlanması probleminin ən çətin tərəfi ondan ibarətdir ki, anomaliyanın normal kimi təsnif edilməsinin dəyəri

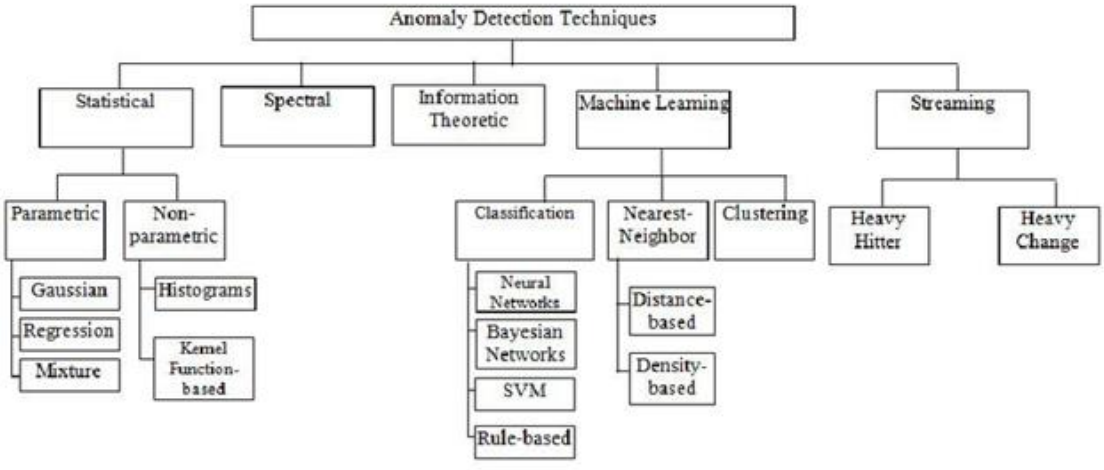
çox yüksək ola bilər. Tibbi anomaliyaların aşkarlanması üçün qabaqcıl tibbi görüntüləmə texnologiyaları və ya digər qabaqcıl tibbi cihazlar tələb olunur. Bununla belə, bu texnologiyaların əldə edilməsi baha başa gələ bilər (Varun Chandola, Arindam Banerjee & Vipin Kumar, 2009).

Kompüter şəbəkəsində anomaliyalar şəbəkə trafikində yüksək yanlış həyəcan siqnalları, emal ediləcək əhəmiyyətli miqdarda məlumat, trafik məlumatlarını vaxt tələb edən paket səviyyəsində və real vaxt rejimində təhlil etmək, yeni tip hücumları aşkarlamaq və s. kimi problemlərin aradan qaldırılması məqsədi ilə dissertasiya işində süni intellek, maşın təlimi əsasında anomaliyaların aşkarlanması üçün ansambl klassifikatoru sistemi təklif edilmişdir.

# II FƏSİL. ŞƏBƏKƏ TRAFİKİNDƏ ANOMALİYALARIN AŞKARLANMASI METODLARININ ANALİZİ

## 2.1. Anomaliyaların aşkarlanması sahəsində tədqiqatların müasir vəziyyətinin analizi

Proqram-aparat təminatı və şəbəkə topologiyalarının mürəkkəbləşməsi və təkmilləşməsi ilə kiberhücumlar davamlı olaraq inkişaf edir. Şəbəkəni zərərli kiberhücumlardan qorumaq üçün müdaxilənin aşkarlanması sistemlərinin işlənilməsi çox vacibdir. Şəbəkə anomaliyalarının aşkarlanması uzun bir tarixə malikdir, onun əsasları 1980-ci illərdə kompüter sistemlərinə müdaxilə edənlərin qarşısını almaq vasitəsi kimi qoyulmuşdur. Dorothy E. Denning, kompüter təhlükəsizliyi sahəsində ilklərdən biri, bir çox anomaliya aşkarlama üsulları üçün əsas olacaq müdaxilənin aşkarlanması üzərində işləmişdir (D. E. Denning, 1987). Bu iş Qauss paylama məlumatlarını klassifikasiya etmək üçün 1930-cu illərdə istehsal sənayesində istifadə edilən daha əvvəlki anomaliyaların aşkarlanması üsulları üzərində qurulmuşdur. Sonrakı onilliklər ərzində tədqiqatçılar müxtəlif növ kənar göstəriciləri aşkar etməyin daha səmərəli yollarını tapmaq üçün maşın öyrənmə üsullarından istifadə edərək anomaliyaların aşkarlanması prosesini avtomatlaşdırdılar. Anomaliyaların aşkarlanmasının ümumi metodlarına statistik, maşın öyrənmə, dərin öyrənmə, neyron şəbəkəsi və s. alqoritmlər daxildir (Şəkil 2.1).




**Şək. 2.1** Anomaliya aşkarlanma metodları (Baddar, Sherenaz & Merlo, Alessio & Migliardi, Mauro, 2014).

Statistik anomaliya aşkarlanması məlumatların əksəriyyətindən əhəmiyyətli dərəcədə kənara çıxan nadir və ya qeyri-adi məlumat nöqtələrini müəyyən etmək üçün istifadə edilən bir üsuldur. Bu, məlumat nümunələrini baza səviyyəsi ilə müqayisə etməyi və nümunələr bazadan kənarda olduğu müəyyən ediləndə tədbirlər görməyi əhatə edir. Statistik metodlar parametrik və qeyri-parametrik olaraq iki yerə bölünürlər.

Laura Feinstein, Dan Schnackenberg və başqalarının "Statistical Approaches to DDoS Attack Detection and Response" (L. Feinstein, D. Schnackenberg, R. Balupari, D. Kindred, 2003) məqaləsində seçilmiş paket atributlarının entropiya və tezlik üzrə çeşidlənmiş paylamalarını hesablamaqla DDoS hücumlarını müəyyən etmək üsullarını təqdim ediblər. DDoS hücumları anomaliyaları seçilmiş paket atributlarının xüsusiyyətlərində görsənir. Aşkarlama dəqiqliyi və performansı İnternetin nüvəsindəki nöqtələrdən tutmuş kənar şəbəkə daxilində olanlara qədər müxtəlif şəbəkə mühitlərindən canlı trafik izlərindən istifadə etməklə təhlil edilir. Nəticələr göstərir ki, bu üsullar cari hücumlara qarşı effektiv ola bilər və daha gizli hücumların aşkarlanmasının təkmilləşdirilməsi üçün istiqamətlər təklif edir.

Maşın öyrənməsinin (Machine Learning) anomaliyalraın aşkarlanması sahəsində önəmli rolu var. A. B. Nassif, M. A. Talib və başqalarının araşdırmalarına görə (A. B. Nassif, M. A. Talib, Q. Nasir, F. M. Dakalbab, 2021) 2000-2020-ci illərdə bu sahədə yazılan 290 elmi məqalədə anomaliya aşkarlanmasının 43 fərqli tətbiqi müzakirə edilmişdir. Bununla birlikdə, anomaliya aşkarlanmasında 29 fərqli ML modeli istifadə edilmişdir.

Hiroshi Hamamoto, Fernando Carvalho və başqaları (Anderson Hiroshi Hamamoto, Luiz Fernando Carvalho, Lucas Dias Hiera Sampaio, Taufik Abrão, Mario Lemes Proença, 2018) şəbəkə anomaliyalarının aşkarlanması üçün Genetik Alqoritm və Qeyri-səlis Məntiqi birləşdirən sxem tətbiq ediblər. Genetik Alqoritm, "Flow" Analizindən istifadə edərək şəbəkə trafikinin müəyyən bir zaman intervalı üçün davranışını proqnozlaşdırıb şəbəkə seqmentinin rəqəmsal imzasını yaratmaq üçün istifadə olunub. Bundan əlavə, nümunənin anomaliya olub-olmamasına qərar vermək üçün qeyri-səlis məntiq sxemi tətbiq edilmişdir. Təklif olunan anomaliya aşkarlama sistemi şəbəkə problemlərini avtonom şəkildə müəyyən edir. Yanaşmanın

real şəbəkə trafik axınlarında tətbiqindən əldə edilən nəticələr 96,53% dəqiqliyə və 0,56% yanlış positiv göstəriciyə nail olmuşdur.

Son tədqiqatlar göstərir ki, dərin öyrənmə (Deep Learning) üsulları müxtəlif sahələrdə, məsələn, təbii dil emalı (Natural Language Processing), nitqin tanınması, kompüter görmə, kibertəhlükəsizlikdə və s. tətbiq edilir. Çünki onlar balanssız siniflər (class imbalance) və qeyri-xətti xassələri olan yüksək ölçülü verilənləri idarə etmək imkanlarına malikdirlər. Bir sinif digərindən əhəmiyyətli dərəcədə daha çox nümunəyə (instance) sahibdirsə, verilənlər bazasında sinif balanssızlığı var. Balanssız verilənlər bazası ən azı ikili sinif və ya çox sinifli ola bilər. Bununla birlikdə, dayaq vektor maşını (Support Vector Machine), Naive Bayes (NB), qərar ağacı (Decision Tree), təsadüfi meşə (Random Forest) və daha çox şərti maşın öyrənməsi (ML) alqoritmləri şəbəkə anomaliyalarının aşkarlanması üçün bir çox tədqiqatçılar tərəfindən təklif edilmişdir. Lakin əsas məhdudiyyət ondan ibarətdir ki, bu alqoritmlər modelin performansını qiymətləndirmək üçün yalnız yaxşı balanslaşdırılmış şəbəkə trafiki məlumatlarıyla işləyir (Hooshmand, M.K., Hosahalli, D., 2022). Standart alqoritmlərin əksəriyyəti balanssız verilənlər toplusunda əksəriyyət sinfinə qarşı qərəzli nəticəyə gətirib çıxarır, çünki onlar çoxluq təşkil edən sinfi azlıqdan üstün tuturlar (Batista, G.E., Prati, R.C., Monard, M.C., 2004; Singh, A., Purohit, A., 2015).

Dünyada şəbəkə trafikinin həcmi getdikcə artmaqdadır. 2023-cü ildə qlobal internet “bandwith” 1,217 Tbps olmuşdur. Bu sürətlə böyüyən rəqəmsal dünyada “Big Data” və Dərin öyrənmə məlumat elminin (Data science) əsas diqqət mərkəzidir. “Big Data” ənənəvi alətlərdən istifadə etməklə idarə edilməsi və analizi çətin olan böyük miqdarda rəqəmsal xam məlumatların toplusudur. Rəqəmsal məlumatlar müxtəlif formalarda, formatlarda və ölçülərdə eksponensiyal olaraq böyüdüyünə görə, bu böyük həcmli məlumatı qurumun ehtiyaclarına uyğun idarə etmək çox vacibdir. Ənənəvi məlumatların emal üsulları böyük həcmli məlumatların işlənməsi üçün bir sıra məhdudiyyətlərə malikdir. Məlumatların həcmi artdıqca dərin öyrənmə ənənəvi maşın öyrənməsi metodlarını üstələməkdədir. Son illərdə dərin öyrənməyə əsaslanan anomaliyanın aşkarlanması alqoritmləri daha çox istifadə olunmağa başlanmışdır. Dərin öyrənmə yüksək ölçülü, müvəqqəti, məkan və qrafik məlumatları daxil olmaqla,

mürəkkəb məlumatların ifadəli təsvirlərini öyrənmək üçün son zamanlarda öyrənmə alqoritmlərini xeyli inkişaf etdirmişdir. Anomaliyaların aşkarlanmasında dərin öyrənmə neyron şəbəkələrdən istifadə edir. Məlumat dəstini öyrənmək üçün dərin neyron şəbəkələri hesablama qrafiki ilə təmsil oluna bilən xətti və qeyri-xətti funksiyaların mürəkkəb birləşmələrindən yararlanır. Dərin neyron şəbəkələri giriş, gizli, və çıxış qatlar və aktivləşdirmə funksiyalarından ibarətdir. Aktivləşdirmə funksiyaları qovşağın çıxışını onun fərdi girişlərinə və onların çəkilərinə əsasən hesablayır. Bu funksiyalar xətti və ya qeyri-xətti ola bilərlər. Sigmoid, ReLU (Rectified Linear Unit) və s. geniş istifadə olunan aktivləşdirmə funksiyalarındandır. Neyron şəbəkələrində qat müxtəlif yollarla üst-üstə düzülmüş neyronlar toplusudur. Tez-tez istifadə olunan qatlara tam əlaqəli, konvolyusiya və birləşən və təkrarlanan qatlar daxildir. Bu qatlar geniş istifadə olunan müxtəlif neyron şəbəkələri qurmaq üçün istifadə edilə bilər. Məsələn, tam əlaqəli qatlar çoxqatlı perseptron (Multilayer Perceptron Networks) şəbəkələrini təşkil edir; konvolyusiya və birləşən qatların müxtəlif qrupları konvolyusiya neyron şəbəkələrini (Convolutional Neural Networks) xarakterizə edir; və təkrarlanan neyron şəbəkələri (Recurrent Neural Networks), qapalı təkrarlanan vahidlər (Gated Recurrent Units) və uzun qısamüddətli yaddaş (Long Short Term Memory) təkrarlanan qatlar üzərində qurulur (Ian Goodfellow, Yoshua Bengio, Aaron Courville, 2016).

Deep learning sahəsində həm xarici, həm də yerli tədqiqatçıların gördüyü işlər diqqət çəkir. DDoS hücumlarının aşkarlanmasında klassifikasiya və klasterləşmə alqoritmlərini istifadə edərək maşın öyrənmə metodları hazırlanmışdır. Bu sahəylə məşqul olan yerli tədqiqatçılar Yadigar İmamverdiyev və Fərqanə Abdullayeva DoS hücumların aşkarlanması üçün deep learning metodu olan Qauss-Bernoulli tipli məhdudlaşdırılmış Boltzmann maşınından (Restricted Boltzmann Machine - RBM) istifadə ediblər. DoS hücumunun aşkarlanmasının dəqiqliyini artırmaq üçün RBM-in görünən və gizli təbəqələri arasına yeddi əlavə təbəqə əlavə olunub. Təklif olunan çoxqatlı dərin Qauss-Bernoulli tipli RBM-dən daha yüksək dəqiqlik əldə edilmişdir (Imamverdiyev Y, Abdullayeva F, 2018).

## 2.2 Statistik metodların analizi

Şəbəkə trafikində anomaliyaların statistik metodlarla aşkarlanması zamanı subyektlərin fəaliyyətləri müşahidə edilir və onların yerinə yetirdikləri fəaliyyətlərə uyğun olaraq profillər yaradılır. Profillər əsasən fəaliyyət intensivliyini, audit qeydlərinin paylanmasını, kateqoriyaları (kateqoriyalar üzrə fəaliyyətin paylanmasını) xarakterizə edən meyarları özündə ehtiva edir. Əksər hallarda, hər istifadəçi üçün iki profil yaradılmış olur: cari profil və qeydə alınmış profil. Sistem/şəbəkə üzrə fəaliyyət yerinə yetirildikcə cari profillər yenilənir və dövri olaraq anomaliya qiyməti hesablanılır. Əgər anomaliya dəyəri təyin edilmiş limitdən yüksək olarsa, bu zaman xəbərdarlıq verilir. Anomaliyaların statistik metodlarla aşkarlanması zamanı təhlükəsizlik zəiflikləri və ya kiberhücumların özləri haqqında məlumatlı olmağa ehtiyac duyulmur. Bunun nəticəsində də, "zero day" hücumlarının və ya ən son növ kiberhücumların aşkarlanması mümkündür. Əlavə olaraq, statistik metodlar DoS kimi uzun zaman aralığında baş verən kiberhücumlara məxsus göstəricilərin vaxtında aşkarlanmasında da effektiv yanaşmadır. Məsələn, portların skanlanması zamanı statistik metodun tətbiqi ilə normal şəbəkə trafikinə nisbətdə yüksək dərəcədə anomaliyanın ehtiva olunduğu şəbəkə trafikini qeydə alınmış olacaq. Apardığımız tədqiqatlar göstərdi ki, sadalanan üstünlüklərlə yanaşı, anomaliyaların statistik metodlarla aşkarlanması zamanı bəzi çatışmazlıqlar da mövcuddur. Bacarıqlı kibercinayətkar anormal davranışı normal davranış qəbul edəcək şəkildə anomaliyaların aşkarlanması üçün istifadə edilən statistik metodları təlim etdirə bilər. Bundan əlavə, "false-positive" və "false-negative" halların azaldılması üçün balanslaşdırılmış limitlər təyin etmək də olduqca çətindir və statistik metodlar dəqiq statistik paylanmalar əsasında işlədiyindən bütün davranışları sırf statistik metodlardan istifadə etməklə modelləşdirmək mümkün deyil (Animesh Patcha & Jung-Min Park, 2007; P. Garcı´a-Teodoroa, J. Dı´az-Verdejoa, G. Macia´-Ferna´ndeza & E. Va´zquezb, 2008).

Ümumilikdə, statistik metodlar parametrik və qeyri-parametrik olmaqla iki kateqoriyaya bölünür (Monowar H Bhuyan, Dhruba Kumar Bhattacharyya & Jugal K Kalita, 2017; Nauman Shahid, Ijaz Haider Naqvi & Saad Bin Qaisar, 2015). Qeyri-parametrik yanaşmalar verilmiş məlumatın statistik xüsusiyyətləri barədə heç bir

təxmin ehtiva etmirlər. Onlar modeli işlədikcə yaradırlar və məlumat nöqtələrinə effektiv şəkildə adaptasiya etmək üçün məlumatın mürəkkəbliyini aradan qaldırmağa çalışırlar. Ən sadə qeyri-parametrik statistik yanaşmalardan biri histoqram alətlərindən istifadə etməkdir ki, onlar məlumatın cədvəlləşdirilmiş tezliklərini qrafik şəklində göstərirlər (ARi Vasudevan, E Harshini & S Selvakumar, 2011). Şəbəkə trafikində anomaliyaların aşkarlanması məqsədilə istifadə edilən müdaxilə və nasazlıqların aşkarlanması sistemlərində normal histoqram qurulur və sonra yeni sınaq nöqtələri müəyyənləşdirilir ki, əgər onlar normal histoqramın daxilində olmasalar, anomaliya halları kimi qəbul edilirlər.

Statistik metodlardan istifadə etməklə müdaxilənin aşkarlanması üsullarına (Nong Ye & Qiang Chen, 2001)-də ehtiva edilən anomaliyaların aşkarlanması üçün qurulmuş xi-kvadrat nəzəriyyəsini nümunə çəkə bilərik. Bu yanaşmaya əsasən informasiya sistemində normal hadisələrin profili yaradılır. Bu yanaşmanın əsas ideyası böyük həcmə malik trafikin içərisindən həm anomaliyaları, həm də müdaxilələri aşkar etməkdir. Xi-kvadrat test statistikasına əsaslanan məsafə meyarları (Sheenam & Abhinav Bhandari, 2016)-da daha ətraflı təsvir edilən formulaya əsaslanır:

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - E_i)^2}{E_i} \qquad (2.1)$$

$\mathbf{X_i}$ = i-ci dəyişənin müşahidə dəyəri,

$\mathbf{E_i}$ = i-ci dəyişənin gözlənilən dəyəri,

**n** = dəyişənlərin sayı.

(Krügel C, Toth T & Kirda E, 2002) şəbəkə trafikinində anomaliyaları, daha dəqiq formada R2L və U2R kimi nadir hücumları aşkar etmək üçün statistik emal meyarları təklif etmişdir. Bu istiqamətdə, müxtəlif xidmət sorğularının eyni xüsusiyyətlərini avtomatik olaraq axtarmağa imkan verən bir metrik hazırlanmışdır. Sorğunun anomaliya qiyməti aşağıdakı üç əsas xüsusiyyət əsasında hesablanır:

- sorğunun növü;
- sorğunun uzunluğu;
- mübadilə edilən məlumatın paylanması.

Şəbəkə administratoru anormal sorğular üçün xəbərdarlıq bildirişi vermək üçün tələb olunan limiti müəyyən edir. Anomaliya qiyməti (bundan sonra – AS) aşağıdakı tənlikdəki kimi hesablanır və burada mübadilə edilən məlumatın paylanması digər xüsusiyyətlərə nisbətən daha çox ağırlıq ehtiva edir.

$$AS = 0.3 \times AS_{növ} + 0.3 \times AS_{uzunluq} + 0.4 \times AS_{yük} \qquad (2.2)$$

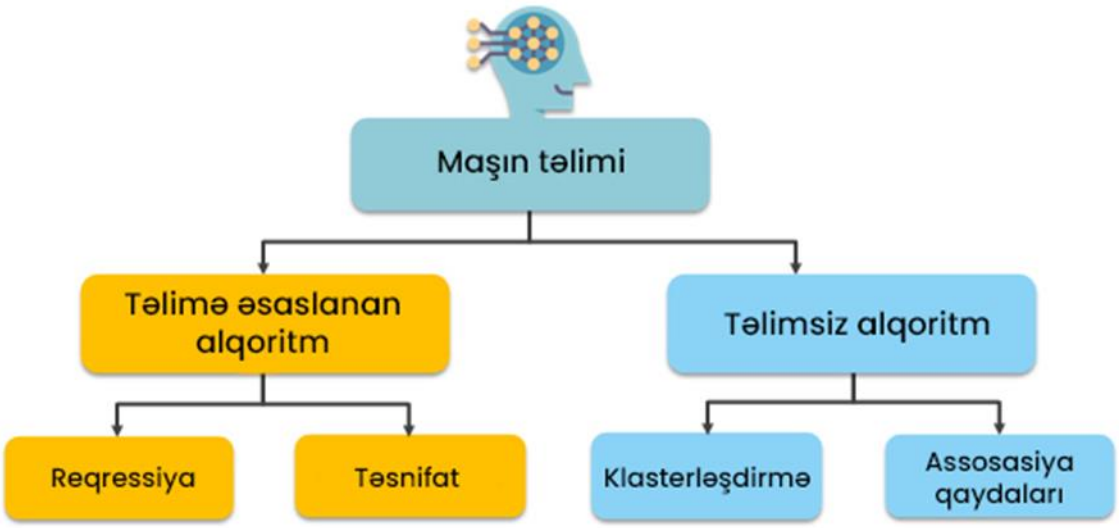
Bir çox tədqiqat işləri ilə yaxından tanışlığımız şəbəkə trafikində anomaliyaların aşkarlanması məqsədilə istifadə edilən müdaxilə və nasazlıqların aşkarlanması sistemlərinin qurulmasında da statistik metodlardan istifadə edildiyini deməyimizə əsas verir. Belə ki, paket filterlənməsindən və neyron şəbəkələrdən istifadə edərək müdaxilələri müəyyən edən sistem (J. M. Bonifacio, 1996) tərəfindən təqdim edilmişdir. Proqram sisteminə qarşı anomaliya və naməlum müdaxilələri aşkar etmək üçün neyron şəbəkələrin istifadəsi ilk dəfə Anderson tərəfindən (J. P. Anderson, 1980) öyrənilmişdir. (Caberera, João & Ravichandran, B. & Mehra, Raman, 2000) tədqiqat işində Kolmoqorov-Smirnov statistikası hücumları araşdıraraq DoS-ı modelləşdirmək və aşkar etmək üçün istifadə edilmişdir. (Manikopoulos, C., & Papavassiliou, S, 2002) tədqiqat işində Konstantin və Saymon İyerarxik müdaxilənin aşkarlanması (HIDE) sisteminin prototipi təklif etmişdir ki, bu sistem şəbəkə hücumlarını və nasazlıqları aşkar etmək üçün statistik metod və neyron şəbəkə təsnifatından istifadə edir.

## 2.3 Maşın təlimi metdolarının analizi

Maşın Öyrənməsi, kompüter sistemlərinin müəyyən bir tapşırığı yerinə yetirmək üçün məlumatların təhlili və nümunənin tanınması kimi üsullardan istifadə edərək təcrübədən öyrənmək imkanı qazandığı süni intellekt sahəsidir. Maşın öyrənmə alqoritmləri böyük həcmli məlumatlara əsaslanan modellər qurur və gələcək məlumatlar əsasında proqnozlar və ya qərarlar vermək üçün bu modellərdən istifadə edə bilər.

Maşın öyrənməsi müxtəlif sahələrdə müxtəlif məqsədlər üçün istifadə edilə bilər [Nikunj C. Oza a, Kagan Tumer]. Nikunj Oza və Kagan Tumer maşın təliminin real dünyada tətbiqlərini izah etmiş, təbii dil emalı, səhiyyə, kompüter şəbəkələri, hərbi sənayədə necə istifadə olunduğuna dair nümunələr vermişdir.

"Təlimli" və "Təlimsiz" maşın öyrənməsi süni intellekt sahələrində istifadə olunan üç əsas təlim paradiqmasıdır (Şəkil 2.2).




**Şək. 2.2** Maşın təliminin qrafik təsviri (Rəşad Səfərov, 2024).

***Təlimə əsaslanan öyrənmə:***

Təlimə əsaslanan öyrənmə (supervised learning) etiketli məlumat dəstlərindən istifadə etməklə həyata keçirilir. Bu məlumat dəstlərində hər bir məlumat nümunəsi giriş və hədəf çıxış dəyərlərindən ibarətdir. Alqoritm bu verilənlər toplusu üzərində öyrədilir və giriş dəyərlərindən hədəf nəticələri proqnozlaşdırmağı öyrənir. Məsələn, təsvirdəki obyektləri müəyyən etmək və ya e-poçt mesajlarını spam və ya qeyri-spam kimi təsnif etmək kimi tapşırıqlar nəzarət edilən öyrənmə problemlərinin

nümunələridir. Təsnifat və reqressiya kimi problemlər nəzarət edilən təlimin əhatə dairəsinə düşür.

***Təlimsiz Öyrənmə:***

Təlimsiz öyrənmə (unsupervised learning) etiketlənməmiş məlumat dəstlərindən istifadə etməklə həyata keçirilir. Bu məlumat dəstlərində məlumat nümunələri yalnız giriş dəyərlərindən ibarətdir. Alqoritm bu məlumat dəstində nümunələri, qrupları və ya strukturları kəşf etməyə çalışır. Məsələn, marketinq şirkəti müştəriləri müəyyən xüsusiyyətlərə görə qruplara bölməklə müxtəlif marketinq strategiyaları hazırlaya bilər. Klasterləşmə və ölçülərin azaldılması (dimension reduction) kimi problemlər nəzarətsiz öyrənmənin əhatə dairəsinə düşür.

### 2.3.1 Təlimə əsaslanan alqoritmlər

Şəbəkədə baş verən anomaliyaların tətqiqi üçün geniş istifadə olunan alqoritmlərdən biri təlimli alqoritmlərdir (və ya nəzarət olunan öyrənmə) [Hacirəhimova M.Ş., Yusifova L.R. 2022]. Nəzarət olunan öyrənmə (Supervised learning) iki əsas tipə bölünür:

**Reqressiya:** Bir və ya bir neçə müstəqil dəyişən kombinasiyası əsasında nəticə dəyişənini proqnozlaşdırmaq məqsədi daşıyır. Nümunələrə xətti reqressiya, çoxhədli reqressiya və dəstək vektor reqresiyası daxildir.

**Təsnifat(classification):** Məlumat nöqtələrini müəyyən bir kateqoriyaya təyin etmək məqsədi daşıyır. Nümunələrə dəstək vektor maşınları (SVM), qərar ağacları, k-NN (k yaxın qonşular) və süni neyron şəbəkələri daxildir.

***Dayaq vektor maşınları (Support Vector Machine- SVM)***

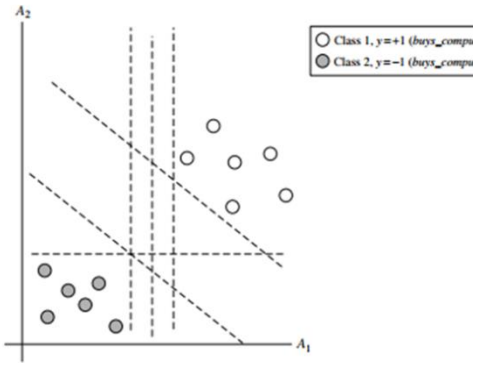
Təlimə əsaslanan alqoritmlər üçün istifadə olunan texnologiyalardan biri məhz dayaq vektor maşınlarıdır. Vladimir Vapnik və həmkarları tərəfindən AT&T Bell Laboratories-də hazırlanan SVM ən yaxşı proqnoz metodlarından sayılır. SVM 2 sinifə aid olan bir data set-də məlumatları təsnifat etmək üçün ortada qalan verilənlərə yeni təlimatlar göndərir, bundan sonra həmin nöqtələrin hansı sinifə daha yaxın olduğuna baxılır və o sinifə aid edilir. SVM nisbətən bəsit texnika sayılır və böyük olmayan data set-lərdə, siniflər arasında böyük fərq olan situasiyalarda daha səmərəli hesab olunur.

Dayaq Vektor Maşınları (SVM) kompüter şəbəkələrində anomaliyaların aşkarlanması üçün effektiv şəkildə istifadə edilə bilər. SVM xüsusilə təsnifat tapşırıqlarında istifadə olunan maşın öyrənmə alqoritmidir.

Hər biri iki kateqoriyadan birinə aid olduğu qeyd olunan bir sıra təlim nümunələri nəzərə alınmaqla, SVM təlim alqoritmi bu və ya digər kateqoriyaya yeni nümunələr təyin edən bir model qurur və onu qeyri-mümkün ikili xətti təsnifləndiriciyə çevirir. SVM, iki kateqoriya arasındakı boşluğu artırmaq üçün yerdəki nöqtələrə təlim nümunələri göstərir. Daha sonra yeni nümunələr eyni məkanda göstərilir və klasterin hansı tərəfinə düşdüklərinə görə kateqoriyaya aid olduğu proqnozlaşdırılır. SVM həm ədədi proqnozlaşdırma, həm də təsnifat üçün istifadə edilə bilər və əlyazma, rəqəm

tanıma, obyekt tanıma, natiq identifikasiyası və zaman seriyası proqnozlaşdırma meyarları da daxil olmaqla bir sıra sahələrə tətbiq edilmişdir.

Fərz edilir ki, təlim toplusu elementlərindən ibarətdir, burada - əlamətlər vektoru, isə ona uyğun sinif nişanıdır. Elə hiper-müstəvi tapmaq lazımdır ki, o və nöqtələrini ayırsın və təlim çoxluğunun ən yaxın nöqtələrindən maksimal məsafədə keçsin. əlamətlər fəzasını siniflərə bölən hiper-müstəvini təsvir edir. Burada, hiper-müstəvinin normal vektorudur. Əgər vektorunun ilə skalyar hasili -nin icazə verilən qiymətindən böyükdürsə, onda yeni nöqtə birinci kateqoriyaya, azdırsa, $w \cdot x_i < b \Rightarrow y_i = -1$, ikinciyə aiddir [Carlos A. Catania, Facundo Bromberg, Carlos Garcнa Garino, 2012]. Vizuallaşdırmağı asanlaşdırmaq üçün şəkil 2.3-də göstərildiyi kimi A1 və A2 iki giriş atributuna əsaslanan bir nümunəyə baxaq. Şəkil 2.3 ölçülü məlumatların xətti olaraq ayrıldığını görə bilərik, çünki +1 sinifinin bütün sərhədlərini -1 sinifinin bütün sərhədlərindən ayırmaq üçün düz bir xətt çəkilə bilər.




**Şək. 2.3** SVM metodunun işləmə qrafiki. (Carlos A. 2012)

Siniflər arasında dəqiq bir fərq olduğu zaman nisbətən yaxşı işləməyi, yüksək ölçülü fəzalarda və ölçülərin sayının nümunə sayından daha çox olduğu hallarda daha təsirli olmağı, yaddaşın nisbətən səmərəli olmağı SVM-in əsas üstünlükləridir.

Lakin SVM alqoritmi böyük məlumat dəstləri üçün uyğun deyil. Məlumat dəsti daha çox anomal olduqda, yəni hədəf sinifləri üst-üstə düşdüyündə SVM çox yaxşı nəticə vermir. Hər bir məlumat nöqtəsi üçün xüsusiyyətlərin sayının təlim məlumat nümunələrinin sayından çox olması hallarında, SVM çatışmazlıq göstərir. Dəstək vektor təsnifatçısı məlumat nöqtələri qoyaraq işlədiyindən, təsnifat hiperplanının yuxarı və aşağı hissələrində təsnifat üçün ehtimal olunan bir izah yoxdur [Jiawei H., Micheline K.]. Tədqiqatçılar [Catania C.A., Facundo B., Carlos Garcнa Garino 2012;

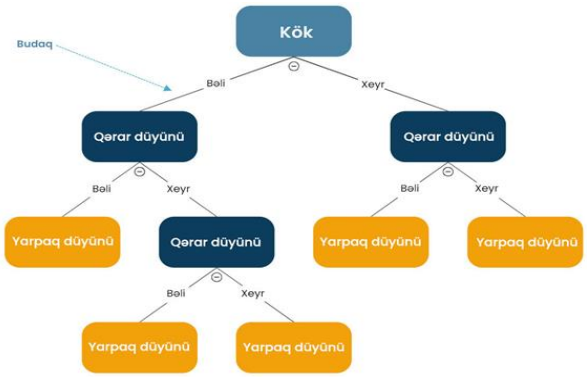
Erfani S.M., Rajasegarar S, Karunasekera S, Leckie C.201; Agarwal B., Mittal N.2012]-də böyük ölçülü şəbəkə trafiki verilənlərində anomaliyaların aşkarlanmasında SVM alqoritmini tətbiq etmişlər.

Tədqiqatlardan da göründüyü kimi anomaliya trafikini düzgün müəyyənləşdirmək üçün əvvəlcədən təyin edilmiş bir qaydanı təyin etmək çox çətin olduğundan normal və anomaliya trafiki arasında böyük fərq bilinmir. [Agarwal B., Mittal N.2012]-də anomaliya trafikini aşkarlamaq üçün həm şəbəkə xüsusiyyətləri entropiyasının, həm də dəstək vektor maşınının (SVM) birləşməsindən ibarət olan hibrid yanaşma təklif olunmuşdur. [Zhang J., Zulkernine M., and Haque A. 2008]-də anomaliyaları aşkarlamaq üçün hibrid metodlar təklif olunmuşdur. Agarwal B. və həmkarları şəbəkə xüsusiyyətləri entropiyasının və dəstək vektor maşını (Support Vector Machine - SVM) alqoritmlərinin birləşməsindən ibarət hibrid metod təklif etmişlər. Nəticədə entropiya əsaslı aşkarlama metodu şəbəkədəki anomaliyaları SVM əsaslı aşkarlama sistemindən daha yaxşı müəyyən etdiyi sübut edilmişdir.

***Qərar ağacı (decision tree)***

1970-ci illərin sonu və 1980-ci illərin əvvəllərində, maşın öyrənmə tədqiqatçısı J. Ross Quinlan ID3 (İterativ Dichotomiser) kimi tanınan bir qərar ağacı alqoritmini təklif etmişdir.

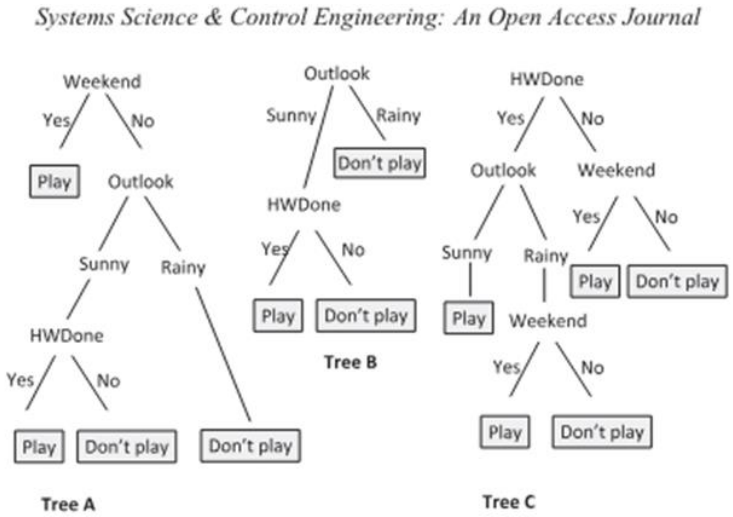
Şəkil 2.4-də göründüyü kimi, qərar ağacı budaqları olmayan kök ilə başlayır. Kök qovşağından çıxan budaqlar daha sonra qərar qovşaqları kimi tanınan daxili qovşaqlara (internal node) qidalanır. Mövcud xüsusiyyətlərə əsasən, hər iki qovşaq növü yarpaq qovşaqları (leaf node) və ya terminal qovşaqları ilə işarələnən homogen alt çoxluqlar yaratmaq üçün qiymətləndirmələr aparır. Yarpaq qovşaqları verilənlər bazasında bütün mümkün nəticələri təmsil edir:




**Şək. 2.4** Qərar ağacı alqoritminin vizual təsviri. (Rəşad Səfərov, 2024).

***Təsadüfi meşələr (Random Forests)***

Alqoritm ilk dəfə L. Breyman tərəfindən verilmişdir. Təsadüfi meşələr ən çox istifadə olunan texnologiyalardan biridir, səmərəliliyinə görə həm təsnifat həm də reqresiyyada istifadə oluna bilir. Təsnifat zamanı yekunda sinif, reqressiya zamanı isə rəqəm çıxarır. Əsas ideyası isə bir neçə qərar ağacının topluluğuna əsaslanır. Bəzi data set-lərdə qərar ağacları səmərəli olmadığından Random Forest texnologiyasına gedilir. Hər bir ağac tam təsadüfi şəkildə yaradılır, bu da bir başa ağacların nəticələrinin bir-birindən fərqlənməsinə gətirib çıxarır. Təsnifat zamanı, meşəni təşkil edən ağacların hər birinin nəticəsi alınıb, sonda onların səsverməsi yekun RF-in nəticəsi sayılır. Reqressiya zamanı isə ağacların nəticələrinin ortalaması yekun nəticə sayılır. Alqoritmin arxitekturası şəkil 2.5-də verilmişdir [Khaled Fawagreh, Mohamed Medhat Gaber, Eyad Elyan 2014].




**Şək. 2.5** Təsadüfi meşələr alqoritminin arxitekturası. (Khaled Fawagreh, 2014).

Atribut seçmə metodu, verilmiş qapaqları sinifə görə "ən yaxşı" şəkildə ayıran bir atributu seçmək üçün bir evristik tərif edir. Bu prosedurda məlumat alma və ya Gini indeksi kimi bir atribut seçmə metrikası istifadə olunur.

$$\text{Gini(D)} = 1 - \sum_{i=1}^{m} p_i^2 \qquad (2.3)$$

Bir ağacın ciddi şəkildə ikili olub olmaması ümumiyyətlə atribut seçimi ölçüsü ilə müəyyən edilir. Bəzi xüsusiyyət seçmə tədbirləri, məsələn, Gini indeksi, yaranan ağacın ikili olmasına məcbur edir. Digərləri, məlumat qazanma kimi, çox yollu

bölmələrə (yəni bir qovşaqdan iki və ya daha çox budağın yetişdirilməsinə) imkan vermir.

Breiman təqdim etdiyi original variantda xəta payı (error rate) korelasiyadan və gücdən asılıdır [L. Breiman, 2001]. Belə ki, korelasiya artan zaman ağaclar arasında olan xəta payı da artır. Aşağı xəta payına sahib ağac isə güclü sinifləndiricidir. Güc artdıqda isə xəta payı düşür. Nəticə olaraq, Bernard, Heutte, və Adam 2010-cu ildə belə nəticəyə gəldilər ki, güc artması və korelasiyanın azalması xəta payının minimuma düşməsinə səbəb olur.

Təsadüfi Meşə kompüter şəbəkələrində anomaliyaları aşkar etmək üçün effektiv şəkildə istifadə edilə bilər. Təsadüfi Meşə bir çox qərar ağaclarının birləşməsindən ibarət ansambl öyrənmə üsuludur. Kompüter şəbəkələrində anomaliyaları aşkar etmək üçün Random Forest-in necə istifadə olunacağına dair bir yanaşma göstərmək mümkündür.

Tətbiqi:

Qiymətləndirilmiş Random Forest modeli real vaxt şəbəkə trafikində həyata keçirilə bilər. Şəbəkə trafikini təhlil edərək, model anormal qarşılıqlı əlaqələri aşkarlaya və müvafiq tədbirlər görə bilər (məsələn, xəbərdarlıq göndərmək və ya avtomatik təcrid etmək) [Cafer M.Y. 2020, Jiong Z., Mohammad Z., Anwar H., 2008]. [Zhang J., Zulkernine M., and Haque A, 2008]-də Zhang və həmkarları IDS-lərin problemlərini həll etmək üçün, qaydalara əsaslanan (ruled-based) xüsusilə yeni müdaxilələri aşkar etmək üçün təlimsiz təsadüfi meşələr (random forests) maşın təlimi alqoritmi tətbiq etmişdir.

***K ən yaxın qonşu (KNN)***

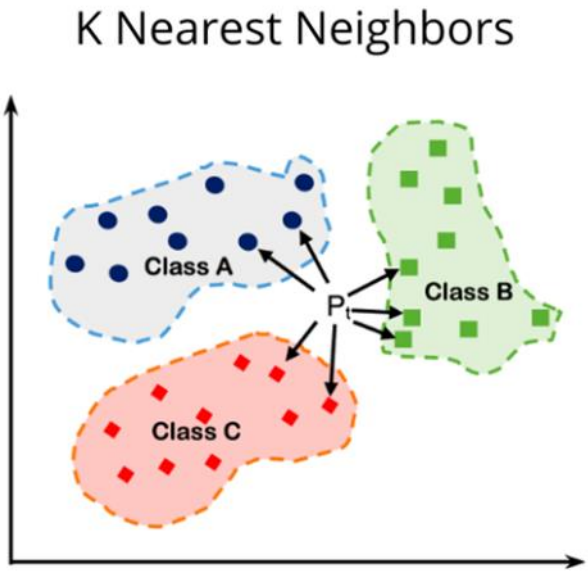
KNN nəzarət edilən öyrənmə alqoritmidir, yəni verilənlər bazasındakı nümunələrin onlara təyin olunmuş etiketləri olmalıdır. KNN haqqında bilmək üçün daha iki vacib nüans var. Birincisi, KNN qeyri-parametrik alqoritmdir. Bu o deməkdir ki, model istifadə edilərkən verilənlər bazası haqqında heç bir fərziyyə aparılmır. Bunun əvəzinə, model tamamilə təqdim olunan məlumatlardan qurulur. İkincisi, KNN-dən istifadə edərkən məlumat toplusunun təlim və test dəstlərinə bölünməsi yoxdur. KNN təlim və test dəsti arasında ümumiləşdirmə aparmır, buna

görə də modeldən proqnoz vermək istənildikdə bütün təlim məlumatları da istifadə olunur.

Alqoritm beş addımdan ibarətdir:

1. Əvvəlcə K qiyməti müəyyən edilir.
2. Digər obyektlərdən hədəf obyektə qədər olan Evklid məsafələri hesablanır.
3. Məsafələr sıralanır və minimum məsafəyə əsasən ən yaxın qonşular tapılır.
4. Ən yaxın qonşu kateqoriyaları cəmlənir.
5. Ən uyğun qonşu kateqoriyası seçilir.

Şəkil 2.6-də qrafik təsvirə nəzər yetirək:




**Şək. 2.6** KNN alqoritminin qrafik təsviri. (Rəşad Səfərov, 2024).

Aşağıdakı düstur vasitəsi ilə Evklid məsafəsini hesablamaq mümkündür:

$$Fark_e = \sqrt{(X_i - X_{yeni})^2 + (Y_i - Y_{yeni})^2} \tag{2.4}$$

Nəzarət olunmayan şəbəkə anomaliyalarının aşkarlanması üzrə ilk işlərdən biri [L. Portnoy, E. Eskin, S. Stolfo 2001]-də təqdim edilmişdir, burada təlim verilənlər bazasında kənar göstəriciləri aşkar etmək üçün tək bağlantılı klasterləşdirmə alqoritminin variantından istifadə olunur. Normal nümunələr kənar nümunələrdən ayrıldıqdan sonra, nəzarət edilən aşkarlama modelini qurmaq üçün normal məlumatların qruplarından istifadə olunur. [E. Eskin, A. Arnold, M. Prerau, L. Portnoy, S. Stolfo, 2002]-də, Eskin və başqaları anomaliyaların aşkarlanması üçün həndəsi çərçivə

təqdim etdi. Nümunələr F xüsusiyyət məkanından yeni F0 xüsusiyyət məkanına uyğunlaşdırılır və anomaliyalar F0-ın seyrək bölgələrində yerləşən nümunələri axtararaq aşkarlanır. Üç fərqli təsnifat texnikasından istifadə olunur: klasterləşdirmə alqoritmi, k-NN alqoritmi və SVM əsaslı. Təcrübələr KDD-UCI verilənlər bazasında həm şəbəkə trafikini ehtiva edən hissədə, həm də ardıcıllıq və ya sistem çağırışlarını ehtiva edən hissədə aparılır.

KNN ən çox istifadə olunan təlimli alqoritrmlərdəndir, iş prinsipi sadə və başadüşülən olduğundan sinifləndirmə və reqressiya problemləri üçün daha praktiklər yanaşmalar təqdim edir.

***Sadəlövh Bayes (Naive Bayes)***

Sadəlövh Bayes klassifikatoru maşın öyrənməsində bir təsnifat alqoritmidir və təlimli alqoritmlərə daxil edilmişdir. Bu alqoritm Thomas Bayes tərəfindən yaradılan Bayes Teoreminə əsaslanır.

Bayes teoremi ehtimal nəzəriyyəsində istifadə olunan Önəmli mövzulardandır. Bu teorem, təsadüfi bir dəyişən üçün ehtimal paylanması içərisində şərti ehtimallarla marjinal ehtimallar arasındakı əlaqəni göstərir. Beləliklə, Bayes teoremi bütün statistiklər üçün məqbul hesab olunan əlaqəni təsvir edir. Bu konsepsiya üçün Bayes qaydaları və ya Bayes qanunu adları da istifadə olunur. Bununla birlikdə, bəzi statistiklər üçün Bayes teoremi xüsusilə fərqli bir əhəmiyyətə malikdir.

Ehtimal nəzəriyyəsində öyrənilən bir hadisə olaraq, B hadisəsi ilə şərtlənən bir A hadisəsi üçün ehtimal dəyəri (yəni B hadisəsi bilinən halda A hadisəsi), B hadisəsi üçün ehtimal dəyəri şərti olaraq A hadisəsinə (yəni A hadisəsi olduğu halda B hadisəsi məlumdur) ehtimal dəyərindən fərqlidir. Bununla birlikdə, bu iki əks şərt arasında çox spesifik bir əlaqə var və bu əlaqəyə ilk dəfə izah edən İngilis statistikisti Tomas Bayesin (1702-1761) adından sonra Bayes teoremi deyilir.

Bayes teoremi stoxastik proses zamanı baş verən təsadüfi bir A hadisəsi ilə başqa bir təsadüfi hadisə B üçün şərti ehtimallar və marjinal ehtimallar arasındakı əlaqədir.

$$P(A|B) = \frac{\mathrm{P(B|A)P(A)}}{\mathrm{P(B)}} \qquad (2.5)$$

Bayes teoremi düsturundakı hər bir terminə xüsusi adlar verilir:

P (A) termini A üçün ilkin ehtimal və ya marjinal ehtimal adlanır.

P (A | B) termini nəzərə alınaraq, B üçün A-nın şərti ehtimalı adlanır.

P (B | A) termini nəzərə alınaraq, A üçün B-nin şərti ehtimalı adlanır.

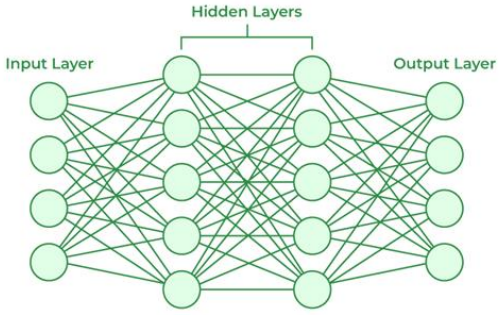
P (B) termini B hadisəsi və ya B-nin hədd ehtimalı üçün "ilkin" ehtimaldır və riyazi rolu normallaşdıran bir sabitdir. Bayes teoreminin çox riyazi olmadan, intuisiyaya əsaslanaraq belə izah edə bilərik: Bayes teoremi, B müşahidə edildiyi təqdirdə A müşahidə ilə bağlı inamların necə yenilənə biləcəyini açıqlayır.

Naive Bayes alqoritmi ümumiyyətlə sadəlik, sürət və yaxşı performans kimi üstünlüklər təklif edir. Bununla belə, müstəqillik fərziyyəsinə və bəzi digər məhdudiyyətlərə görə bəzi hallarda digər alqoritmlərlə müqayisədə zəif çıxış edə bilər. Buna görə də, istifadə olunacağı konkret problem kontekstinə uyğun olaraq qiymətləndirilməlidir.

Tədqiqatlar nəticəsində deyə bilərik ki, sadəlövh Bayes alqoritmi anomaliyaların aşkarlanmasında, monitorinq prosesində istifadə edilir, müəlliflər [Joa˜o B.D. Cabrera, Carlos Gutie´rrez, Raman K. Mehra, Yong Fang, Yunyun Zhang, Cheng Huang, 2019]-da qeyd etdiyimiz nüansları araşdırıb, nəticələrini analiz etmişlər.

***Neyron şəbəkə (Neural Network)***

İnsan beynindəki bir-biri ilə birləşmiş sinir hüceyrələrini (neyronları) modelləşdirən kompüter proqramıdır. 1943-cü ildə yaradılıb, Neyron şəbəkələri kompüterə, insanda olduğu kimi, özünü təlim etməklə şablonları tanımağa imkan verir. İnsan beyni kimi, neyron şəbəkələri də yalnız təxmini nəticələr verir, ancaq onların edə bildiyi işləri başqa növ kompüter proqramlarının heç biri effektiv yerinə yetirə bilmir. Hər bir neyron bir neçə girişə və yalnız bir çıxışa sahib olur.




**Şək. 2.7** Neyron Şəbəkənin qrafik təsviri. (Rəşad Səfərov,2024).

Süni Neyron Şəbəkələri (ANN) çoxlu süni neyron və ya qovşaqlardan ibarət şəbəkədir (Şəkil 2.7). Bu neyronlar giriş məlumatlarını qəbul edir, bu girişləri emal edir və çıxış istehsal edir. Yeni məlumatların təsnifatına başlamazdan əvvəl şəbəkəni dəqiq tənzimləmək və parametrlərini təyin etmək üçün həm təlim, həm də sınaq tələb olunur. Neyron şəbəkələri məlumatların üstünə bir səth yerləşdirməyə çalışır və səthi ayırd etmək üçün kifayət qədər məlumat sıxlığı olmalıdır. Əksər neyron şəbəkələr əsas xüsusiyyətlərə diqqət yetirmək üçün giriş xüsusiyyətlərini avtomatik olaraq azaldır. Bununla birlikdə, xüsusiyyət seçimi və ya daha aşağı ölçülü məlumat proqnozlarından hələ də faydalanırlar [59].

$$f(x)=b+w1\cdot x1+w2\cdot x2+...+wn\cdot xn \quad \text{vəya} \quad f(x)^{\wedge} = \sum\_(i=1)^{n} [\![w\_i \; x\_i ]\!] \tag{2.6}$$

Burada w1, w2...wn -giriş verilənləri x1, x2...xn – çəkiləri göstərir.

Süni neyron şəbəkələri geniş tətbiqlərdə istifadə oluna bilər və müxtəlif məlumat dəstlərində yaxşı performans göstərir. Bununla belə, o, mürəkkəbdir və düzgün öyrədilməli və konfiqurasiya edilməlidir.

Kompüter şəbəkələrində anomaliyaların aşkarlanması üçün ANN-lərin istifadə oluna biləcəyi bəzi ümumi yollar bunlardır:

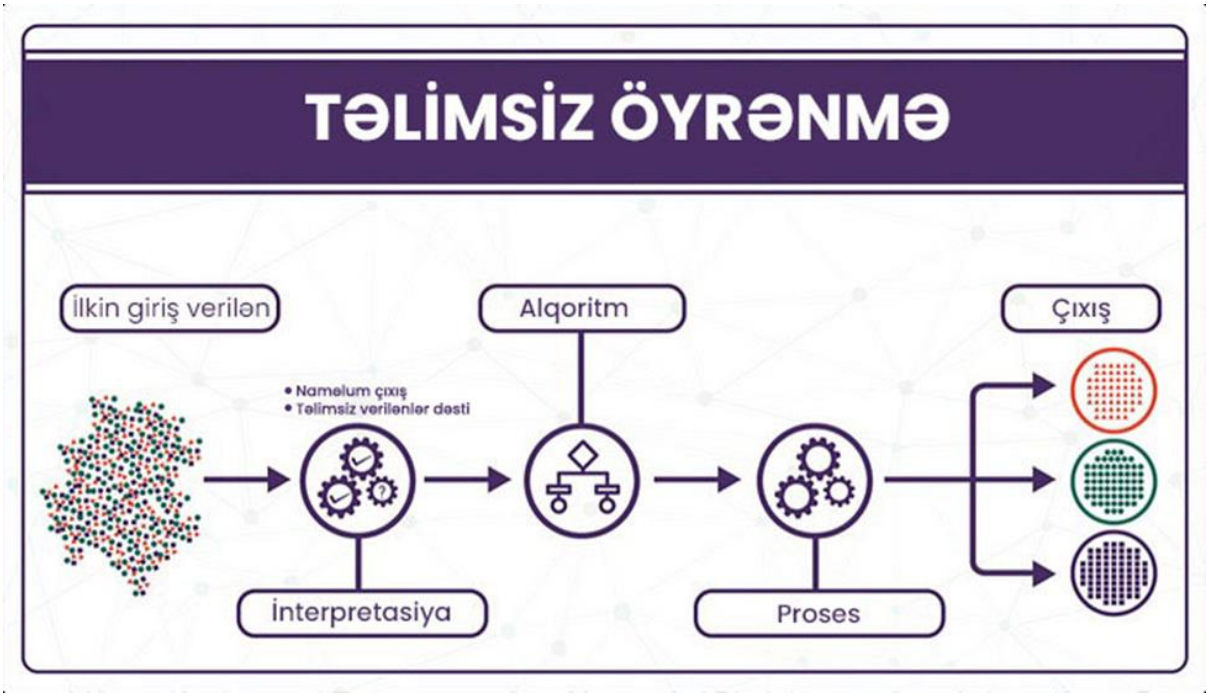
Şəbəkə Trafikinin Təhlili: ANN-lər normal şəbəkə qarşılıqlı əlaqə nümunələri yaratmaq üçün şəbəkə trafiki məlumatlarını təhlil edə və sonra gözlənilməz, anormal şəbəkə trafik nümunələrini aşkar edə bilər. Məsələn, çoxqatlı perseptron (MLP) və ya təkrarlanan neyron şəbəkəsi (RNN) kimi ANN modelləri şəbəkə trafiki məlumatlarını emal etməklə normal və anormal davranışı təsnif edə bilər. [Hacirəhimova M.Ş., Yusifova L.R., 2022]-də Hacırəhimova və Yusifova tədqiqatlarında neyron şəbəkədə MLP-i araşdırıb.

Şəbəkə Firewallları və Təhlükəsizlik insidentlərinin Təhlili: ANN-lər şəbəkə firewalllarından və ya təhlükəsizlik insidentlərindən alınan məlumatları təhlil edərək anomal şəbəkə qarşılıqlı əlaqəsini aşkar edə bilər. Məsələn, şəbəkə qeydlərini və ya təhlükəsizlik insidentlərini ehtiva edən verilənlər dəstlərini emal etməklə o, hücumçu cəhdlərini və ya digər təhlükəsizlik təhdidlərini müəyyən edə bilər [Joa˜o B.D. Cabrera, Carlos Gutie´rrez, Raman K. Mehra, 2008].

Zaman sıralamasının təhlili: ANN şəbəkədəki zaman sıralaması məlumatlarını təhlil edərək normal və anormal davranışı müəyyən edə bilər. Məsələn, RNN və ya avtomatik kodlayıcı kimi ANN modelləri şəbəkə trafiki, server performansı və ya istifadəçi qarşılıqlı əlaqəsi kimi zaman seriyası məlumatlarını təhlil edə və anormal nümunələri aşkar edə bilər [İmamverdiyev Y., Abdullayeva F., 2018]. [İmamverdiyev Y., Abdullayeva F., 2018]-də İmamverdiyev və həmkarı neyron şəbəkənin köməyi ilə DOS hücumunun təyin edilməsi prosesini araşdırıblar.

### 2.3.2 Təlimsiz alqoritmlər

Alqoritm etiketlənməmiş verilən üzərindən öyrənməyə başlayır, əldə olunan nəticələri isə kateqorizasiya edərək, yekun nəticə çıxarır. Yəni, əlimizdə olan böyük verilənlər çoxluğunda verilənlər arasındakı qarışıqlığı həll edib, istədiyimiz şəkildə kateqoriyalaşdıran alqoritm təlimsiz alqoritm adlanır (Şəkil 2.8).




**Şək. 2.8** Təlimsiz öyrənmə alqoritminin strukturu. (Rəşad Səfərov,2024)

Təlimsiz öyrənmənin 2 əsas tipi var:

1. Klasterləşmə (clusstering).
2. Əlaqələndirmə (association)

**Klasterləşdirmə(clusstering)**

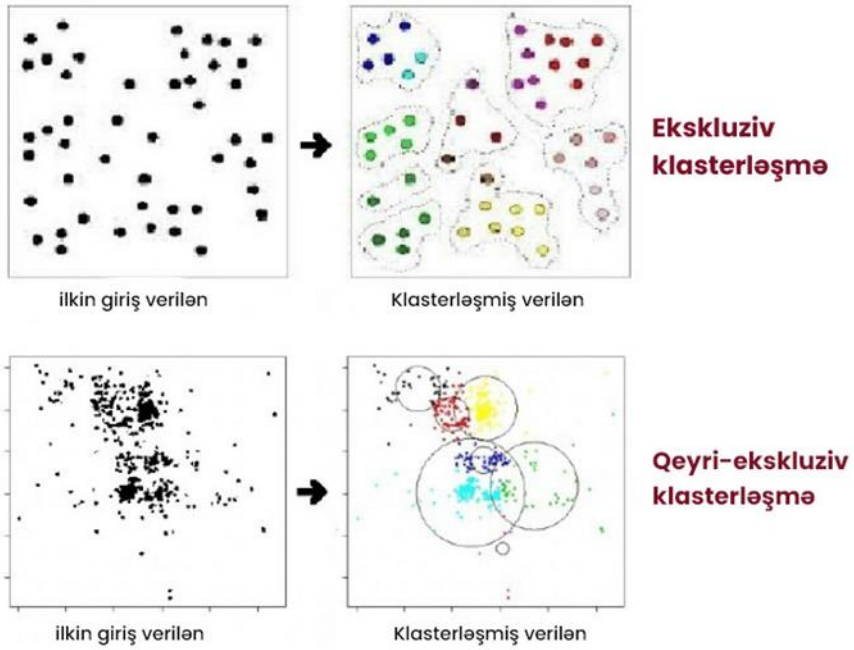
Klasterləşdirmə verilənlərin aralarında olan oxşarlıqlara əsaslanaraq onları qruplaşmasını təmin edir. Təlimli alqoritmlərdən fərqli olaraq ancaq girişlərdən istifadə edilir. Adətən, verilənlər arasında oxşarlıqları tapmaq üçün işlədilir. Praktikada tətbiqi genişdir, real həyata istinad etsək, real mail-lərin spamlardan ayrılması, böyük kompaniyaların istifadəçilərini maraqlarına görə qruplaşdırması və s. kimi nümunələri göstərmək olar.

Klasterləşmə birbaşa verilənlərdən asılı olduğu üçün unikal seçilmiş bir alqoritm yoxdur, daha çox hər situasiya üçün uyğun seçilmiş alqoritmlər çalışdırılır. İstifadə

tipinə görə klasterləşməni 3 qrupa bölünə bilər [Samreen Naeem, Aqib Ali, Sania Anam Muhammad Munawar Ahmed, 2023]:

1. Ekskluziv klasterləşmə
2. İyerarxial klasterləşmə
3. Üst-üstə düşən klasterləşmə (overlapping).

Ekskluziv klasterləşmə, adından da göründüyü kimi, hər bir verilənin yalnız bir klasterə aid olduğunu göstərir. Daha dəqiq başa düşmək üçün aşağıdakı nümunə verilmişdir (Şəkil 2.9):




**Şək. 2.9** Ekskluziv klasterləşmənin qrafik təsviri. (Rəşad Səfərov,2024).

Ekskluziv klasterləşməyə nümunə olaraq, k-means və ya k-medoid alqoritmlərini nümunə göstərmək olar.

K-means alqoritmi nisbətən ən sadə başa düşülən və buna görə də ən çox istifadə olunan alqoritmdir. Bu alqoritmdə məqsəd klasterlərdə olan verilənlərin oxşarlıqlarının maksimum olması və klasterlər arasında bənzərliyin minimum olmasıdır.

Alqoritmin iş prinsipi aşağıdakı ardıcıllıqla gedir [Kaufman, L., & Rousseeuw, P. J. 2005]:

1. Qrup Mərkəzlərinin İlk Dəyərlərinin Seçilməsi: İlk olaraq, qrup mərkəzlərinin sayı (k) təyin edilir və təsadüfi verilən nöqtələrindən seçilir və ya xüsusi bir metodla başlanğıc dəyərləri təyin edilir.

2. Təyinat Mərhələsi: Hər verilən nöqtəsi, onun qrup mərkəzlərinə olan məsafəsi nəzərə alınaraq ən yaxın klaster mərkəzinə təyin edilir. Məsələn, Evklid məsafəsi məsafə ölçüsü kimi tez-tez istifadə edilir.
3. Mərkəzlərin Yenilənməsi: Hər bir qrup üçün yeni mərkəzlər hesablanır. Bu, hər qrupa təyin edilmiş olan verilən nöqtələrinin ortalaması alınaraq edilə bilər.
4. Təkrar Təyinat və Yenilənmə Addımlarının İterasiyası: Təyinat və yenilənmə addımları təkrarlanır. Verilən nöqtələri qruplara təyin olunur, sonra qrup mərkəzləri yenilənir. Bu proses, qrup mərkəzlərinin və verilən nöqtələrinin bir-birinə yaxın olduğu bir vəziyyət əldə edilənə və ya təyin edilmiş bir kriteriyaya çatdıqlarında dayandırılır.
5. Algoritminın Sonlanması: Təyinat və yenilənmə addımları nəticəsində qrup mərkəzlərinin və təyin olunan veri nöqtələrinin mövqeyi artıq dəyişmirsə və ya müəyyən bir iterasiya(təkrarlanma) sayına çatdıqda alqoritm sonlandırılır.

K-means alqoritmində bəzən, klaster sayının öncədən təxmin edilməsi çətin olur, buna baxmayaraq, ən sadə alqoritmdir, istifadəsi rahatdır və ən əsası effektivdir [Kumari, R., Sheetanshu, Singh, M. K., Jha, R., & Singh, N. K., 2016].

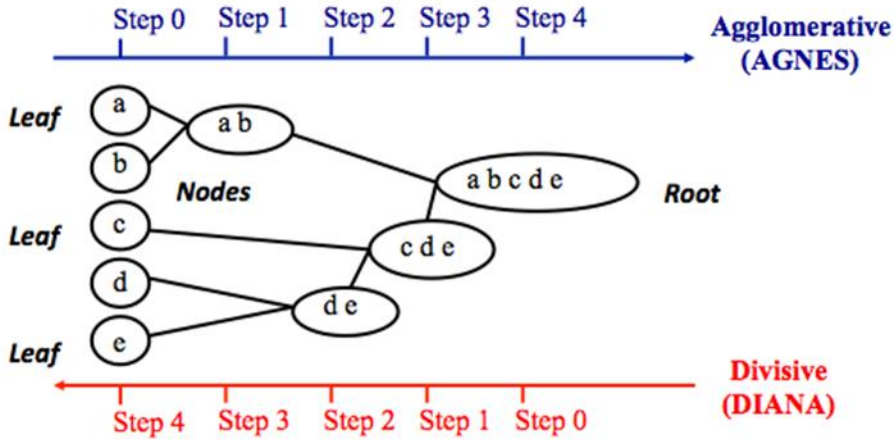
[Münz G., Li S., Carle G, 2007]-də Münz və həmkarı şəbəkə anomaliyalarını aşkarlaya bilən K-means klasterləşmə alqoritminə əsaslanan yanaşma təklif edilmişdir. [Mohammed Hussein Thwaini 2022]-də şəbəkə axın yazılarını ehtiva edən təlim məlumatlarını normal və anomal trafik qruplarına ayrmaq üçün k-means alqoritmi tətbiq edilmişdir. [Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu 2015]-də müəlliflər Apache Spark analitik alətindən istifadə edərək k-means klasterləşmə əsasında kiber-hücumların qarşısını almaq üçün metodologiya təqdim edilmişdir.

İyerarxik klasterləşmə dedikdə isə adından da göründüyü kimi bir növ iyerarxiya yaranır. Nümunə olaraq, canlılar aləmini klasterləşdirərkən burada birhüceyrəlilər və çoxhüceyrəlilər olaraq iki qrupa, daha sonra çoxhüceyrəliləri prostistalar, bitkilər, heyvanlar və göbələklərə bölmək olar. Bu şəkildə sıralı davam edən qruplaşmalar iyerarxik klasterləşmə adlanır.

İyerarxik klasterləşmədə 2 ən çox istifadə edilən alqoritm var: Aqlomerativ və Diviziv.

Aqlomerativ qruplaşma “aşağıdan yuxarıya” məntiqi ilə çalışır. İlk olaraq, verilənlər setindən verilənlərin sayı qədər klaster yaradılır. Sonra bir-biri ilə oxşar olan 2 qrup birləşdirilir, yaradılan yeni qruplar arasında oxşar olanlar birləşdirilərək bu proses davam edir. Öncədən verilmiş hansısa şərt ödəndikdə və ya vahid bir klaster yarandıqda alqoritm sonlanır və nəticə olaraq iyerarxik bir qruplaşma yaranır.

Aqlomerativ qruplaşmanın tərsi isə Diviziv qruplaşmadır (Şəkil 2.10):




**Şək. 2.10** Aqlomerativ və Diviziv klasterləşmənin qrafik təsviri. (S. Ahmadian 2019).

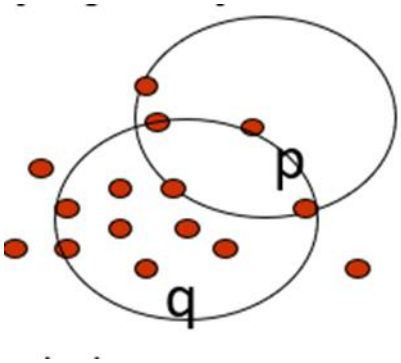
Aqlomerativ qruplaşmanın əksinə olaraq, iyerarxiya “yuxarıdan aşağı” gedir kökdən başlayıb tək bir verilən qalana qədər davam edir. Aqlomerativ klasterləşmə geniş istifadə olunan bir metoddur, lakin bəzi çatışmayan tərəfləri də var. Nümunə olaraq, prosesləri geri qaytarmaq olmur, əgər hər hansı klastering oldusa düzəliş etmək mümkün deyil, böyük verilənlər toplusu üçün əlverişli deyil və istisna verilənlərə qarşı həssasdır. Üstünlükləri isə, tətbiqi rahatdır, qrup sayını təxmin etmək nisbətən daha asandır [S. Ahmadian, A. Epasto, R. Kumar, M. Mahdian, 2019].

***DBSCAN***

(Density-Based Spatial Clustering of Applications with Noise) (Sıxlığa əsaslanan küy tətbiqlərinin məkan klasterləşməsi)- bu nəzarətsiz öyrənmə alqoritmləri yaxın yerləşən (və ya bənzər) təlim məlumatları qruplarının müəyyənləşdirilməsinə çalışır. Elmi ədəbiyyatlarda da çox istinad edilən DBSCAN 1996-cı ildə alman mütəxəssislər, Martin Ester, Hans-Peter Kriegel, Jörg Sander və Xiaowei Xu, tərəfindən təklif edilmiş qeyri parametrik (qeyri-xətti) klasterləşdirmə alqoritmidir [Ester M., Hans-Peter Kr., Jiirg S., Xiaowei Xu,1996]. Bu alqoritm: bəzi

fəzada müəyyən bir nöqtə toplusu üçün bir yerə yığılmış sıx nöqtələri qruplaşdıraraq aşağı sıxlıqlı bölgələrdə tək anomal kənar nöqtələr kimi qeyd edir.

Hər bir obyekt üçün nəzərdən keçirdiyimiz çevrənin radiusunu təyin etmək üçün xüsusi bir parametr €> 0 istifadə olunur. Eps- nöqtənin əhatəsi. P nöqtəsinin Eps əhatəsi NEps (p) kimi işarələnir və NEps(p) = {q ∈D | dist (p, q) ≤ Eps} şəklində müəyyənləşdirilir. Sadə yanaşma, klasterdəki hər nöqtə üçün həmin nöqtənin Eps qonşuluğunda ən azı minimum sayı (MinPts) nöqtəsinin olmasını tələb edə bilər. Küyün olmasına görə klasterdəki hər p nöqtəsi üçün klasterdə q nöqtəsinin olması tələb olunur ki, p q-nın qonşuluğunun içində olsun və NEps (q) ən azı MinPts nöqtəsini ehtiva etsin (Şəkil 2.11).




**Şək. 2.11** DBSCAN alqoritmi. (Rəşad Səfərov,2024)

MinPts = 5

Eps = 1 cm - nöqtənin qonşuluğu

|NEps (q)| ≥ MinPts

Eps - Parametrləşdirilmiş qonşuluqların sabit ölçüsü sayəsində qonşuluğun sıxlığı məhəllədəki obyektlərin sayı ilə ölçülə bilər. Bir qonşuluğun sıx olub olmadığını müəyyən etmək üçün DBSCAN, sıx bölgələr üçün sıxlıq həddini təyin edən başqa bir istifadəçi tərəfindən müəyyən edilmiş MinPts parametri istifadə edir. Ətrafında ən azı MinPts obyekt varsa, bir obyekt əsas obyektdir. Əsas obyektlər sıx bölgələrin sütunlarıdır.

Nəticədə, DBSCAN alqoritmi təbii olaraq məlumat nöqtələrini qonşuluq strukturu və sıxlığı əsasında qruplaşdıra və küyü müəyyən edə bilər. Buna görə də, xüsusilə məlumat dəstindəki klasterlərin sayı naməlum olduqda effektiv klasterləşdirmə alqoritmi kimi istifadə olunur.

DBSCAN alqoritminin üstünlükləri olaraq, Fərqli formalı və ölçülü klasterləri aşkarlaya bilir, fərqli qruplar ilə əhatələnmiş klaster tapır. Daha yüksək aşkarlanmaya davamlıdır, deyə bilirik. Çatışmazlıqları olaraq isə, çox prosessorlu sistemlər üçün paylaşıla bilmir, müxtəlif sıxlığa malik məlumat dəstləri mürəkkəbdir, MinPoints və EPS qruplaşdırma parametrlərinə həssasdır, Eps və MinPts parametrlərini təyin etmək çətindir [Hacirəhimova M.Ş., Yusifova L.R. 2022].

## 2.4 Ansambl metodları: əsasları və alqoritmləri

Maşınla öyrənmə metodlarının qarşısında dayanan əsas problemlərdən biri fərdi modellərin proqnozlarının dəqiqlik səviyyəsinin aşağı olmasıdır. Ansambl metodları özündə bir neçə modeli birləşdirərək anomaliyaların aşkarlanması prosesinin effektivliyini artırır. Bir neçə fərdi detektordan yığılan datanın qarşılaşdırılması daha düzgün qərarların alınmasına və müxtəlif növ anomaliyaların effektiv üsullarla idarə edilməsinə xidmət edir.

Ansambl metodları iki əsas kateqoriyaya ayrılır: ardıcıl və paralel ansambl metodları. Ardıcıl ansambl metodları baza öyrənənləri növbəli şəkildə yaradır. Buna misal olaraq "Adaptive Boosting" (AdaBoost) metodu göstərilə bilər. Baza öyərənənlərin növbəli şəkildə yaradılması onların arasında müəyyən bir asılılığa gətirib çıxarır. Paralel ansambl metodlarında baza öyrənənlər bir-birindən asılı olmadan, paralel şəkildə yaradılır. Bu metodlara misal olaraq "Random Forest" metodunu göstərmək olar.

### 2.4.1. "Voting" ansambl metodu

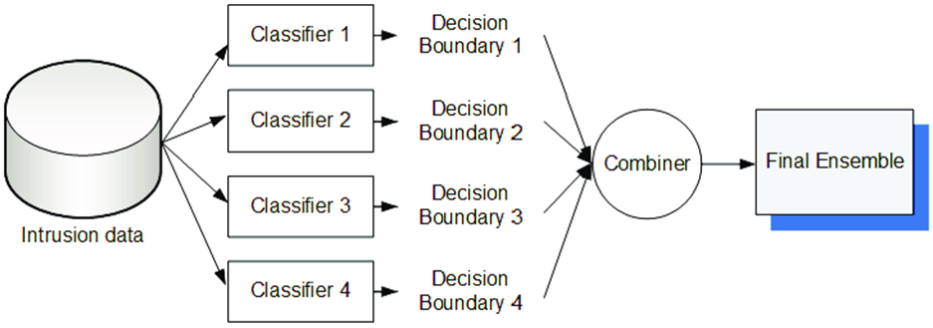
#### 2.4.1.1. Majority voting

"Majority voting" metoduna əsasən, hər bir detektor fərdi olaraq səsvermədə (voting) iştirak edərək verilmiş data nöqtəsinin normal və ya anomaliya olduğuna qərar verir. Bu zaman yekun qərar detektorların daha böyük qismi tərəfindən verilən qərara əsasən müəyyən edilir (Mina Eshak Magdy, Ahmed M. Matter, Saleh Hussin, Doaa Hassan & Shaimaa Ahmed Elsaid, 2023).

İlk addım bir neçə baza modelin eyni dataset vasitəsilə öyrədilməsidir. Həmin baza modelləri müxtəlif maşınla öyrənmə alqoritmlərindən təşkil oluna bilər. Test dataseti vasitəsilə öyrədilmiş baza modelləri yeni datasetlər üzərində proqnozların verilməsi üçün istifadə olunur.

"Majority voting" ansambl metodunda hər bir modelin proqnozu yekun qərara bərabər səviyyədə təsir edir.

"Majority voting" ansambl metodu şəkil 2.12-də vizual olaraq təsvir edilmişdir.

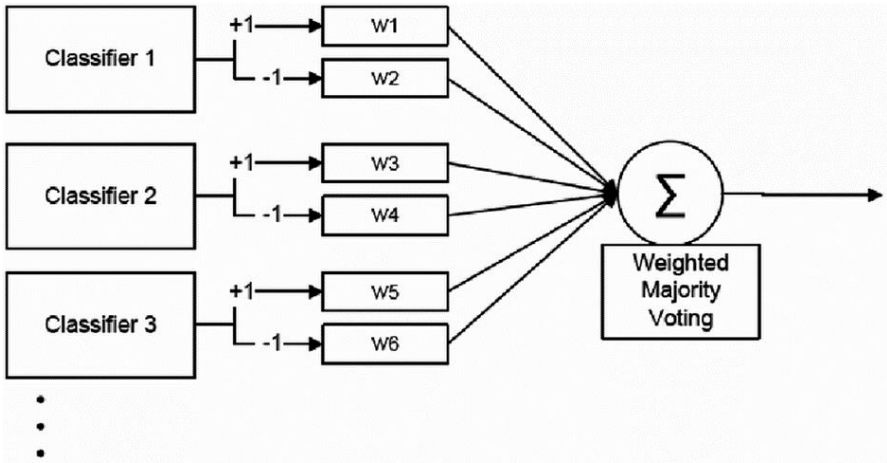



**Şək. 2.12** "Majority voting" metodu ilə yekun qərarın verilməsi (Adhi Tama, Bayu & Rhee, Kyung Hyune, 2017)

### 2.4.1.2 Weighted voting

"Weighted voting" metodunda detektorlar öz etibarlılıq səviyyələrinə əsasən yekun qərara təsir edirlər. Başlanğıcda bir neçə baza modeli eyni dataset üzərində müxtəlif alqoritmlərdən istifadə olunmaqla öyrədilir. Proqnozlaşdırma zamanı hər baza modeli verilən nümunəyə uyğun olaraq öz proqnozunu verir. Effektivliyinə əsasən hər baza modelinə "ağırlıq" (etibarlılıq səviyyəsi) təyin olunur. Müxtəlif etibarlılıq səviyyələrinə məxsus baza modellərin proqnozları toplanılaraq yekun qərar verilir. "Weighted voting" metodu özü də iki növə ayrılır: ən yuxarı etibarlılıq səviyyəsinə malik sinifin seçilməsinə əsaslanan "weighted majority voting" və müxtəlif ağırlıqlı baza modellərinin proqnozlarının ortalamasının alınması ilə həyata keçirilən "Weighted average" (Sergey Sakulin, Alexander Alfimtsev, Konstantin Kvitchenko, Leonid Dobkacz, Yuri Kalgin & Igor Lychkov, 2022).

"Weighted voting" ansambl metodu şəkil 2.13-də vizual olaraq təsvir edilmişdir.




**Şək. 2.13** "Weighted majority voting" metodu ilə yekun qərarın verilməsi (Hulley, Gregory & Marwala, Tshilidzi, 2007)

### 2.4.2 “Bagging” ansambl metodu

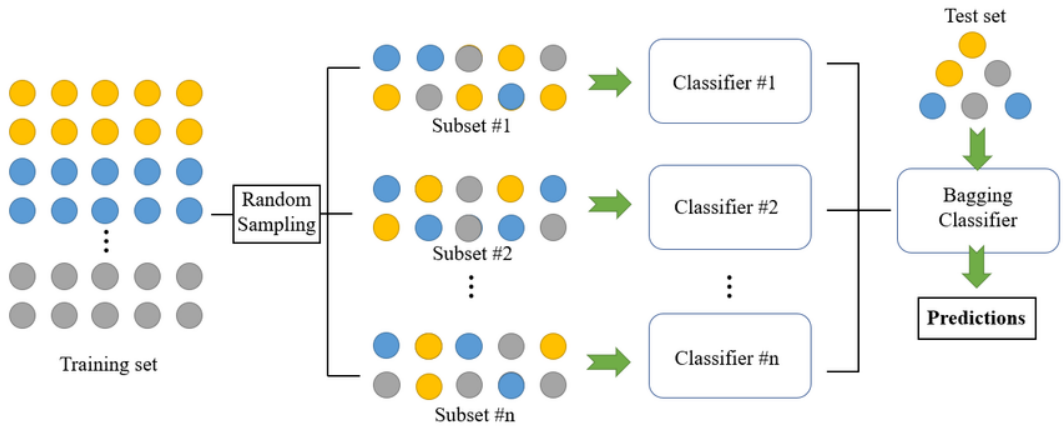
Bu metodda detektorlar datanın təsadüfi seçilmiş hissələri ilə öyrədilir, bununla da detektorların verdiyi qərarların dəqiqliyi artırılır. “Bagging” metodunun tətbiq edilməsi kənaraçıxmaları azaldaraq nəticənin dəqiqliyini artırır və “overfitting”in qarşısını alır.

“Bagging” metodu iki növə ayrılır: “bootstrapping” və “aggregation”. Bootstrapping əvəzləmə proseduruna əsaslanaraq bütün dəstədən nümunələrin götürülməsi metodudur. Əvəzləmə metodu ilə nümunələrin götürülməsi seçim prosedurunda təsadüfiliklə nəticələnir. Prosedurun tamamlanması üçün baza öyrənmə alqoritmi nümunələr üzərində tətbiq olunur.

“Aggregation” metodu proqnozlaşdırılan bütün mümkün nəticələri birləşdirərək tətbiq edilir.

“Bagging” metodunun üstün cəhəti bundan ibarətdir ki, zəif baza öyrənənlər birləşərək bir etibarlı baza öyrənən altında birləşərək daha stabil nəticələr verirlər. Bu, həmçinin hər hansı kənaraçıxmanın da qarşısını alır, modellərdə “overfitting” halını azaldır. Bagging metodunun mənfi cəhətlərindən biri hesablamaların maddi baxımdan əlverişsiz olmasıdır (D. P. Gaikwad & R. C. Thool, 2015).

“Bagging” ansambl metodu şəkil 2.14-də vizual olaraq təsvir edilmişdir.




**Şək. 2.14** “Bagging” metodunun blok-sxemi (Zhang, Hanzhong & Zhou, Ting & Xu, Tianheng & Hu, Honglin, 2023)

### 2.4.3 “Boosting” ansambl metodu

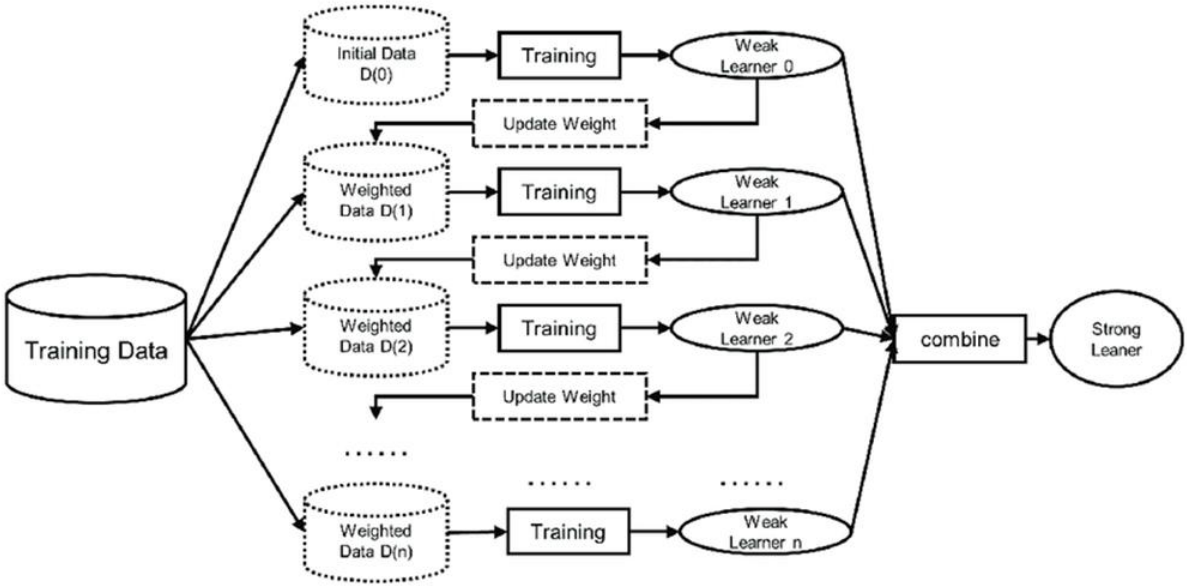
“Boosting” metodu daha əvvəlki proqnoqlaşdırıcılarının səhvləri vasitəsilə öyrənərək gələcək proqnozların dəqiqliyini artırır. Bu metodda bir neçə zəif baza öyrənən birləşərək bir etibarlı baza öyrənən formalaşdırırlar, bu da öz növbəsində

proqnozların dəqiqliyinin əhəmiyyətli səviyyədə artırır. “Boosting” metodunun işləmə prinsipi zəif öyrənənlərin müəyyən bir ardıcıllıqda sıralanmasına əsaslanır. Belə ki, zəif öyrənənlər ardıcıllıqda özündən sonra gələn vasitəsilə öyrənərək daha etibarlı proqnozlaşdırma modeli yaradırlar.

“Boosting” metodunun bir neçə növü var (məsələn, gradient boosting, Adaptive Boosting, XGBoost).

“Gradient boosting” proqnozlaşdırıcıları ansambla ardıcıllıqla əlavə edir. Öncəki proqnozlaşdırıcılar öz davamçılarının səhvlərini düzəldərək modelin dəqiqliyini artırırlar (M. K. Islam, P. Hridi, M. S. Hossain & H. S. Narman, 2020).

“Boosting” ansambl metodu şəkil 2.15-də vizual olaraq təsvir edilmişdir.



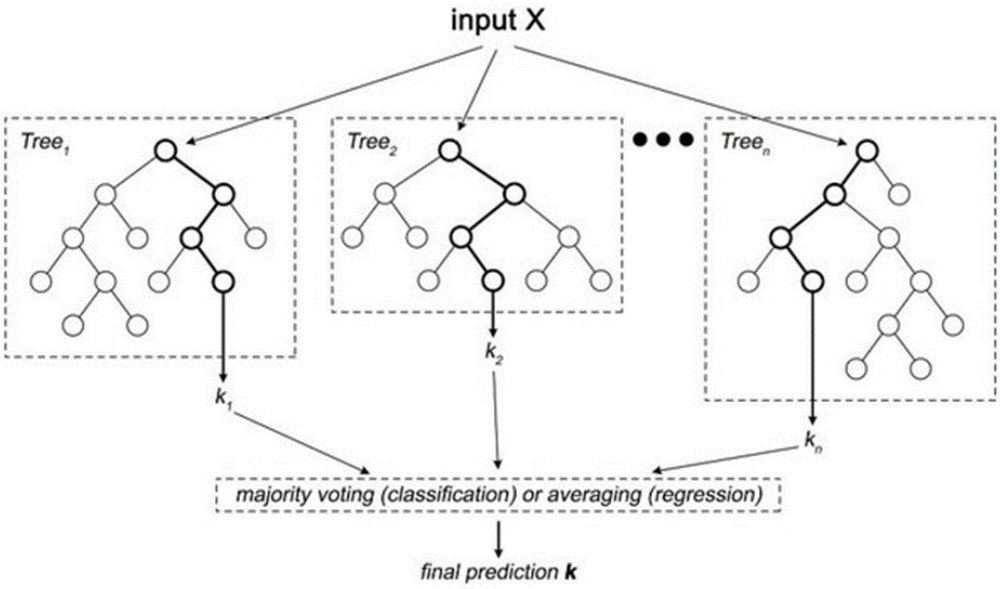

**Şək. 2.15** “Boosting” metodunun blok-sxemi (Xu, Wenqing & Ning, Like & Luo, Yong, 2020)

### 2.4.4 “Random forest” ansambl metodu

Leo Breiman tərəfindən bir sıra məqalə və texniki hesabatlarda (L. Breiman, 1996, 2000, 2001, 2004) təsadüfi parametrlər əsasında böyüyən ağac strukturuna malik ansambl metodlarında klassifikasiya və reqressiya dəqiqliyinin əhəmiyyətli ölçüdə artırıla bilməsi ehtiva olunmuşdur. Yekun proqnozlar isə ansambl üzərindən cəmlənilərək əldə edilir. Ansamblın əsasını ağac strukturlu proqnozçular təşkil edir və hər bir ağac təsadüfi metod əsasında qurulduğundan bu metod “random forest” adlanır.

Breiman'in yanaşmasına görə, kolleksiyadakı hər bir ağac, ilk növbədə, hər düyündə təsadüfi olaraq bölünmək üçün kiçik bir qrup giriş koordinatları seçilməklə və ikincisi, bu giriş koordinatları əsasında təlim toplusunda ən yaxşı bölünmənin hesablanması ilə formalaşdırılır. Ağac, maksimum ölçüyə qədər, budanmadan, CART metodologiyası (L. Breiman, J.H. Friedman, R.A. Olshen & C.J. Stone, 1984) istifadə edərək böyüdülür. Bu randomizasiya sxemi, hər dəfə yeni fərdi ağac yetişdirildikdə təlim məlumatlarını əvəz etməklə yenidən formalaşdırılaraq "bagging" metodu ilə qarışdırılır (Gerard Biau, 2012).

"Random forest" ansambl metodu şəkil 2.16-da vizual olaraq təsvir edilmişdir.




**Şək. 2.16** "Random forest" ansambl metodunun blok-sxemi (Zucco, Adrian Gabriel, 2017)

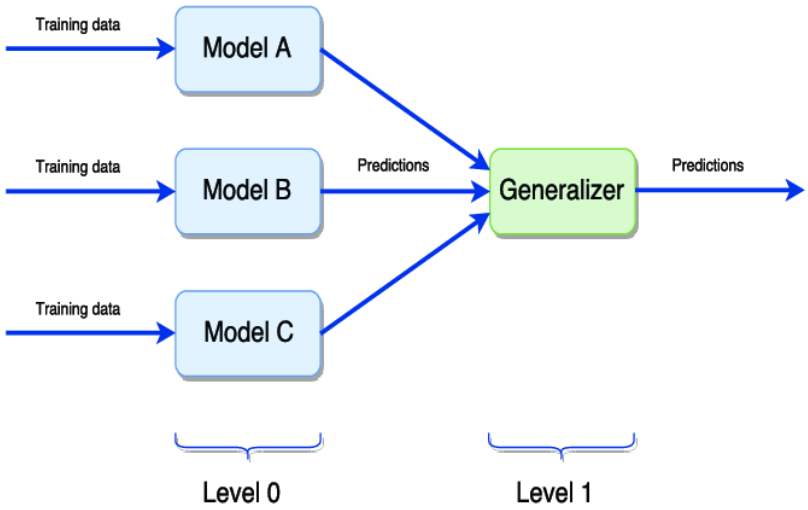
### 2.4.5. "Stacking" ansambl metodu

"Stacking" ansambl modeli, bir neçə maşın öyrənmə modelinin proqnozlarını birləşdirərək ümumi performansı yaxşılaşdıran bir ansambl öyrənmə alqoritmidir. "Stacking" ansambl metodunun əsas fikri bir sıra əsas modellərin ilk növbədə eyni məlumatlar əsasında təlim edilməsidir. Daha sonra bu əsas modellərin proqnozları daha yaxşı proqnoz verməyi öyrənən daha yüksək səviyyəli bir meta-modelin təlim edilməsi üçün istifadə olunur. "Stacking" ansambl metodu həm reqressiya, həm də klassifikasiya problemləri üçün istifadə edilə bilər. Bu, fərdi modellərin performansını adətən yaxşılaşdıra bilən güclü bir ansambl öyrənmə alqoritmidir

(Smitha Rajagopal, Poornima Panduranga Kundapur & Katiganere Siddaramappa Hareesha, 2020).

"Stacking" ansambl modeli bir neçə müxtəlif model növünü birləşdirdiyindən, bir model növündən daha yaxşı performans sərgiləyir və fərdi modellərin güclü və zəif tərəflərindən yararlana bilir. Lakin bu metodun mənfi cəhətləri hesablamaların maddi baxımdan əlverişsiz olması və tətbiqinin çətinliyidir.

"Stacking" ansambl metodu şəkil 2.17-də vizual olaraq təsvir edilmişdir.




**Şək. 2.17** "Stacking" metoduna dair nümunə blok-sxem (Divina, Federico & Gilson, Aude & Gómez-Vela, Francisco & Garcia Torres, Miguel & Torres, José, 2018)

# III FƏSİL. ŞƏBƏKƏ TRAFİKİNDƏ ANOMALİYALARIN AŞKARLANMASI ÜÇÜN MAŞIN TƏLİMİ ALQORİTMLƏRİNİN REALİZASİYASI

## 3.1. Data-set və alqoritmlərin seçilməsi

Araşdırdığımız tədqiqat işlərinə əsaslanaraq bu nəticəyə gəldik ki, 2007-ci ildən bəri, (T. Subbulakshmi, K. BalaKrishnan, S. M. Shalinie, D. AnandKumar, V. GanapathiSubramanian & K. Kannathal, 2011), (A. M. R. K.Munivara Prasad & K. Rao, 2014), (CAIDA UCSD, 2018), (DARPA, 2018), (A. H. Carson Brown, Alex Cowperthwaite & A. Somayaji, 2009), (K. J. Singh & T. De, 2015) və (S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang & F. Tang) kimi tədqiqatçılar DoS/DDoS data-set inkişaf etdirməyə çalışıblar. Lakin tamamlanmamış trafik, anonimləşdirilmiş məlumatlar və köhnəlmiş hücum ssenariləri kimi çatışmazlıq və problemlərə görə, hələ də tədqiqatçılar öz təklif olunan aşkarlama və müdafiə modellərini sınaqdan keçirmək və qiymətləndirmək üçün əhatəli və etibarlı data-set tapmaqda çətinlik çəkirlər. Əksər tədqiqat işlərində təsvir edilmiş çətinliklərlə rastlaşmamağımız üçün 1999-cu ildə ABŞ-ın DARPA (American Defense Advanced Research Projects Association) tərəfindən yaradılmış KDD-99 data-setinə əsaslanan və həmin data-setdəki təkrarlanma, klassifikasiya dərəcəsinin yüksək olması kimi boşluqları özündə ehtiva etməyən daha təkmilləşmiş versiya olan NSL-KDD data-setindən istifadə etməyə qərar verdik (M. Tavallaee, E. Bagheri, W. Lu & A. Ghorbani, 2009). Təlim dəstində izafi yazılar, test dəstində təkrarlanan yazıların olmaması, data-setdə məlum hücumların təlim verilənlər bazasında, yeni hücumların isə test data-setində yerləşdirilməsi (onlar təlim data-setində yoxdur) tədqiqatçılar tərəfindən NSL-KDD bazasının üstünlükləri kimi qeyd edilir (Ramiz Alıguliyev & Makrufa Sh. Hajirahimova, 2019; Y.N. Imamverdiyev & L.V. Sukhostat, 2017).

Şəbəkə trafikində anomaliyaları aşkarlamaq üçün NSL-KDD məlumat bazasının “train.arff” və “test.arff” fayllarından istifadə olunmuşdur. NSL-KDD-nin təlim məlumat bazasında 125.973, test bazasında isə 22.544 yazı nümunəsi vardır. Hər bir yazı 42 atributdan ibarətdir. Sonuncu atribut hər bir yazıya "anomaly", ya da "normal" vəziyyət olaraq etiketlənmişdir.

NSL-KDD verilənlər toplusunda hücumlar dörd kateqoriyaya bölünmüşdür:

- DoS - təcavüzkar qanuni istifadəçilərə göstərilən xidmətləri həddindən artıq yükləyir;
- U2R - təcavüzkar istifadəçi hesabından inzibatçının hesabına giriş əldə etməyə çalışır;
- R2L - hücum edən kompüterdə müəyyən boşluqlardan istifadə edərək lokal istifadəçi hesabına giriş əldə etməyə cəhd edir;
- Probing Attack (Probe) - informasiya təhlükəsizliyini pozmaq məqsədilə kompüter şəbəkəsi haqqında informasiya toplamağa cəhd edilən hücum.

Anna, Kris və digərlərinin tədqiqatlarına (Jurek A, Bi Y, Wu S & Nugent C, 2014) əsaslanaraq geniş istifadə edilən alqoritmlərdən istifadəyə qərar verdik. "Naive Bayes", "Logistic Regression" və "Random Forest" alqoritmləri əsasında "Bagging" klassifikator ansamblını tətbiq etdik.

## 3.2. WEKA proqram platformasında şəbəkə trafiki verilənlərində anomaliyaları aşkarlamaq üçün klassifikator ansamblı alqoritminin tətbiqi

WEKA (Waikato Environment for Knowledge Analysis) proqram platforması Məqalədə şəbəkə trafiki verilənlərinin analizində maşın təlimi alqoritmlərini yerinə yetirmək üçün WEKA platforması seçilmişdir. WEKA açıq proqram təminatının ilkin versiyası 1993-cü ildə Yeni Zelandiyanın Vaikato universitetində Java proqramlaşdırma dilində yazılmışdır. Bu da onun istənilən kompüter platformada istifadəsinə imkan verir. WEKA tədqiqatçılara ilkin emal alətləri, çoxsaylı klassifikasiya və klasterizasiya, reqresiya metodları təqdim edir və nəticələrin vizuallaşdırılması imkanı verir. Sonrakı illərdə WEKA-nın proqram təminatı inkişaf etdirilmiş, tədqiqatçılara daha geniş imkanlar yaradılmışdır. Eksperimentdə WEKA 3.8.6 versiyası istifadə edilmişdir.

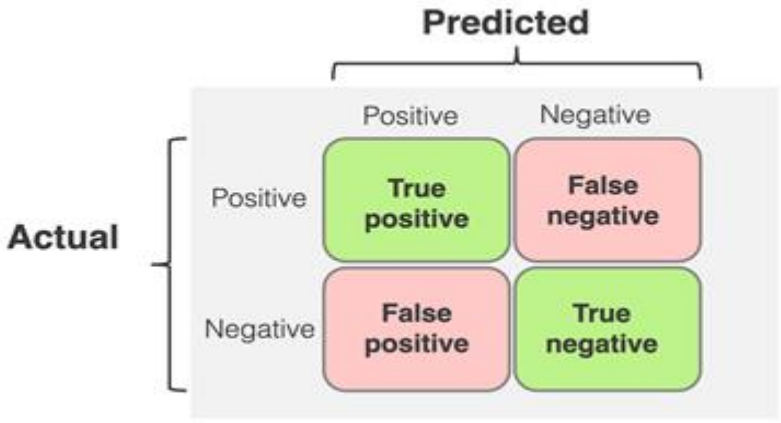
Data bazanı Weka tətbiqində Explorer-də açdıqdan sonra təlim və test setləri yaratmaq üçün orijinal faylı Preprocess bölməsində "RemovePercentage" filteriylə 80/20 nisbətində 2 yerə böldük. Daha sonra bir tərəfdə təlim setini, sağdakı Classify bölməsində isə test setini açırıq.

Maşın təlimində klassifikatorların qiymətləndirilməsində dəqiqlik (precision), tamlıq (recall), yanlış müsbət hallar (false positive rate-FPR), doğru müsbət hallar

(true positive rate - TP), f-ölçü (f-measure), doğruluq (accuracy) metrikalarından istifadə olunmuşdur.

Xətalar matrisi (Confusion Matrix) (Cədvəl 3.1) modelin düzgünlüyünü və dəqiqliyini qiymətləndirmək üçün istifadə edilən ən asan və ən sadə metrikalardan biridir (Fawcett Tom, 2006).

**Cədvəl 3.1** Xətalar matrisi (Kamil Qədirov, 2024).




Xətalar matrisi və onun əsasında hesablanan klassifikasiya alqoritmlərinin aşkarlama göstəriciləri aşağıdakı düsturların köməyi ilə hesablanır (Huang, Jin & Ling, Charles, 2005):

- doğru pozitiv hallar (TPR- true positive rate) - TP/(TP + FN),
- doğru negativ hallar (TNR- true negative rate) - TN/(FP + TN),
- yanlış pozitiv hallar (FPR- false positive rate) - FP/(FP + TN),
- yanlış negativ hallar (FNR- false negative rate) - FN/(TP + FN),
- dəqiqlik (precision) - TP/(TP + FP),
- tamlıq (recall) - TP/(TP + FP),
- f-ölçü (f-measure) - 2 x [(Precision x Recall) / (Precision + Recall)],
- doğruluq (accuracy) - (TP + TN)/(TP + TN + FP + FN).

## 3.3. Eksperimentlərin aparılması: qiymətləndirmə və və nəticələrin müqayisəli interpretasiyası

Şəbəkə trafiki məlumat bazasında DoS hücumların aşkarlanması üçün təklif olunan maşın təlimi ansamblı modelinin yaradılmasında istifadə edilən Random

forests (Şəkil 3.1), Naive Bayes (Şəkil 3.2), Logistic Regression (Şəkil 3.3) və Bagging (Şəkil 3.4) klassifikatorları WEKA tətbiqində test edilmişdir.

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.32 seconds

=== Summary ===

Correctly Classified Instances       18137               80.4516 %
Incorrectly Classified Instances      4407               19.5484 %
Kappa statistic                          0.6199
Mean absolute error                      0.1968
Root mean squared error                  0.3794
Relative absolute error                 38.9795 %
Root relative squared error             74.9868 %
Total Number of Instances            22544

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.973    0.323    0.695      0.973    0.811      0.658    0.959     0.945     normal
               0.677    0.027    0.971      0.677    0.798      0.658    0.959     0.959     anomaly
Weighted Avg.  0.805    0.155    0.852      0.805    0.803      0.658    0.959     0.953

=== Confusion Matrix ===

    a    b   <-- classified as
 9447  264 |    a = normal
 4143 8690 |    b = anomaly
```

**Şək. 3.1** Random Forest alqoritmi ilə klassifikasiyanın nəticəsi (Kamil Qədirov, 2024).

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.16 seconds

=== Summary ===

Correctly Classified Instances       17161               76.1222 %
Incorrectly Classified Instances      5383               23.8778 %
Kappa statistic                          0.5366
Mean absolute error                      0.2386
Root mean squared error                  0.4862
Relative absolute error                 47.2755 %
Root relative squared error             96.0968 %
Total Number of Instances            22544

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.931    0.367    0.657      0.931    0.771      0.572    0.895     0.844     normal
               0.633    0.069    0.924      0.633    0.751      0.572    0.917     0.911     anomaly
Weighted Avg.  0.761    0.197    0.809      0.761    0.759      0.572    0.908     0.882

=== Confusion Matrix ===

    a    b   <-- classified as
 9041  670 |    a = normal
 4713 8120 |    b = anomaly
```

**Şək. 3.2** Naive Bayes alqoritmi ilə klassifikasiyanın nəticəsi (Kamil Qədirov, 2024).

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.06 seconds

=== Summary ===

Correctly Classified Instances       17045               75.6077 %
Incorrectly Classified Instances      5499               24.3923 %
Kappa statistic                          0.5266
Mean absolute error                      0.2437
Root mean squared error                  0.4725
Relative absolute error                 48.2774 %
Root relative squared error             93.388  %
Total Number of Instances            22544

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.926    0.372    0.653      0.926    0.766      0.562    0.777     0.581     normal
               0.628    0.074    0.918      0.628    0.746      0.562    0.777     0.863     anomaly
Weighted Avg.  0.756    0.203    0.804      0.756    0.754      0.562    0.777     0.742

=== Confusion Matrix ===

    a    b   <-- classified as
 8988  723 |    a = normal
 4776 8057 |    b = anomaly
```

**Şək. 3.3** Logistic Regression alqoritmi ilə klassifikasiyanın nəticəsi (Kamil Qədirov, 2024).

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.07 seconds

=== Summary ===

Correctly Classified Instances       18628               82.6295 %
Incorrectly Classified Instances      3916               17.3705 %
Kappa statistic                          0.6552
Mean absolute error                      0.1716
Root mean squared error                  0.3873
Relative absolute error                 34.0024 %
Root relative squared error             76.5577 %
Total Number of Instances            22544

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.911    0.238    0.744      0.911    0.819      0.668    0.928     0.915     normal
               0.762    0.089    0.919      0.762    0.833      0.668    0.928     0.916     anomaly
Weighted Avg.  0.826    0.153    0.843      0.826    0.827      0.668    0.928     0.915

=== Confusion Matrix ===

    a    b   <-- classified as
 8845  866 |    a = normal
 3050 9783 |    b = anomaly
```

**Şək. 3.4** Bagging alqoritmi ilə klassifikasiyanın nəticəsi (Kamil Qədirov, 2024).

Cədvəl 3.2-ə əsasən bütün metrikalar üzrə ən yaxşı nəticəni Bagging alqoritmi göstərmişdir. dəqiqlik metrikası üzrə ən yüksək - 82.6% və yanlış təsnifatlandırma (FP) səhvi üzrə isə ən kiçik – 15.3% nəticə nümayiş etdirmişdir. FP Rate metrikası üzrə ən yuxarı nəticəni isə Logistic Regression alqoritmi göstərmişdir.

**Cədvəl 3.2** Klassifikator nəticələrinin müqayisəsi (Kamil Qədirov, 2024).

| Alqoritm | Accuracy | TPR | FPR | Precision | Recall | F-Measure | MCC | ROC Area |
|---|---|---|---|---|---|---|---|---|
| Random Forest | 0.804 | 0.805 | 0.155 | 0.852 | 0.805 | 0.803 | 0.658 | 0.959 |
| Naive Bayes | 0.761 | 0.761 | 0.197 | 0.809 | 0.761 | 0.759 | 0.572 | 0.908 |
| Logistic Regression | 0.788 | 0.756 | 0.203 | 0.804 | 0.756 | 0.754 | 0.562 | 0.777 |
| Bagging (Ensemble) | 0.826 | 0.826 | 0.153 | 0.843 | 0.826 | 0.827 | 0.668 | 0.928 |

## Nəticə

- Şəbəkə trafikindəki anomaliyaların aşkarlanması, günümüzün dinamik və kompleks kibertəhlükəsizlik mühitində ən əhəmiyyətli məsələlərdən biridir. İnternedən və şəbəkələrdən geniş istifadə nəticəsində qlobal ölçüdə trafikin həcmi və kompleksliyi də təbii olaraq artmaqdadır. Bu, normal trafikdən ayrılan fərqli növ trafiki müəyyən etməyi və potensial təhlükəli fəaliyyətləri tanımağı daha çətinləşdirir. Bu tipli anomaliyaların aşkarlanması problemlərinin həllində elmi ədəbiyyatlarda klassifikator ansamblı kimi tanınan yanaşmalar, son illərdə kibertəhlükəsizlik sahəsinin diqqət mərkəzindədir.
- Ansambllar, fərqli növ trafiki müxtəlif məlumatların qiymətləndirilməsi üçün istifadə edərək, fərqli xüsusiyyətlərə malik alqoritmləri birləşdirir. Beləliklə, normaldan fərqlənən və potensial təhlükəli olaraq təsnif edilə bilən trafiki yüksək dəqiqliklə tanımaq mümkün olur. Lakin, bu texnikaların istifadəsi bəzi məsələləri də gündəmə gətirir. Ən mühüm məsələlərdən biri, klassifikator ansamblılarının işləməsi üçün yüksək miqdarda məlumatın işlənməsi və hesablama gücünün tələb olunmasıdır.
- Müxtəlif klassifikasiya alqoritmlərinin sınağı onu göstərdi ki, şəbəkə trafiki anomaliyalarının - DoS hücumların aşkarlanmasında ansambl əsasında təklif olunmuş yanaşma digər baza təlim klassifikatorları ilə müqayisədə daha yüksək dəqiqlik göstərmişdir. Alınmış nəticələr, 84.3% precision, 82.6% recall və 82.6% doğruluq və 0.928 ROC, DoS hücumların aşkarlanmasında klassifikatorların ansamblının etibarlı həll yolu olduğunu deməyə əsas verir. Təklif olunan yanaşma arzuolunan nəticəni göstərsə də, dərin təlim və optimallaşdırma strategiyalarını və s. tətbiq etməklə daha yüksək nəticələr almaq olar. Eyni zamanda real-vaxt rejimində və real məlumat setindən istifadə etmək də mühüm rol oynayır.

# İSTİFADƏ EDİLMİŞ ƏDƏBİYYAT

A. B. Nassif, M. A. Talib, Q. Nasir and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," in IEEE Access, vol. 9, pp. 78658-78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

A. H. Carson Brown, Alex Cowperthwaite and A. Somayaji, “Analysis of the 1999 darpa/lincoln laboratory ids evaluation data with netadhict,” in Proceedings of the Second IEEE International Conference on Computational Intelligence for Security and Defense Applications, CISDA’09, 2009.

A. M. R. K.Munivara Prasad and K. Rao, “Dos and ddos attacks: Defense, detection and traceback mechanisms - a survey,” Global Journal of Computer Science and Technology, vol. 14, 2014.

Agarwal B., Mittal N., “Hybrid Approach for Detection of Anomaly Network Traffic using Data Mining Techniques”// Procedia Technology 6; 2012; p. 996-1003.

Anderson Hiroshi Hamamoto, Luiz Fernando Carvalho, Lucas Dias Hiera Sampaio, Taufik Abrão, Mario Lemes Proença, Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic, Expert Systems with Applications, Volume 92, 2018, Pages 390-402, ISSN 0957-4174.

Andress, J. “What is Information Security? “The Basics of Information Security, 2014, pp. 1–7.

Animesh Patcha, Jung-Min Park; “An overview of anomaly detection techniques: Existing solutions and latest technological trends”, Elsevier Computer Networks 51 (2007) 3448–3470

Arab Mohammed Shamiulla “Role of Artificial Intelligence in Cyber Security”// International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9 Issue-1, November 2019, pp.4628-4630

ARi Vasudevan, E Harshini, and S Selvakumar. Ssenet-2011: a network intrusion detection system dataset and its comparison with kdd cup 99 dataset. Internet (AH-ICI), 2011 Second Asian Himalayas International Conference, pages 1–5. IEEE, 2011.

B. D. Joao et al., “Statistical Traffic Modeling for Network Intrusion Detection,” Proc. 8th Int’l. Symp. Modeling, Analysis Sim. Comp. Telecommun. Sys., Aug. 2000, pp. 466–73.

Batista, G.E., Prati, R.C., Monard, M.C.: A study of the behavior of

Cabric, M. “Confidentiality, Integrity, and Availability” Corporate Security Management, 2015, pp.185–200

Cafer M. Y. “Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm” // Chaos, Solitons & Fractals Volume 140, November 2020

Carlos A. Catania, Facundo Bromberg, Carlos Garcна Garino An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection Expert Systems with Applications 39 (2012) 1822–1829.

Catania C.A., Facundo B., Carlos Garcна Garino // Expert Systems with Applications vol.39 (2012) pp.1822–1829.

Cybercrime Magazine. https://cybersecurityventures.com/

D. E. Denning, "An Intrusion-Detection Model," in IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp. 222-232, Feb. 1987, doi: 10.1109/TSE.1987.232894.

D. P. Gaikwad and R. C. Thool, "Intrusion Detection System Using Bagging Ensemble Method of Machine Learning," 2015 International Conference on Computing Communication Control and Automation, Pune, India, 2015, pp. 291-295

DARPA 2000 Intrustion Detection Scenario Specific Data Sets, accessed by Jan 2018. https://www.ll.mit.edu/r-d/datasets/ 2000-darpa-intrusion-detection-scenario-specific-data-sets.

Davis David, Zindi: Largest professional network for data scientists in Africa, 2019, https://zindi.africa/blog/introduction-to-anomaly-detection-using-machine-learning-with-a-case-study

E. Eskin, A. Arnold, M. Prerau, L. Portnoy, S. Stolfo, A geometric framework for unsupervised anomaly detection: detecting intrusions in unlabeled data, in: D. Barbara, S. Jajodia (Eds.), Applications of Data Mining in Computer Security, Kluwer, 2002.

Erfani S.M., Rajasegarar S, Karunasekera S, Leckie C. “High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning” //Pattern Recogn. 2016; 58:121–34.

Ester M., Hans-Peter Kr., Jiirg S., Xiaowei Xu “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise” //Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231.

Fawcett Tom 2006; T. Holz 2008; G. Gu, et all, 2006; Ramiz Makrufa-2019.

Firkhan Ali Bin Hamid Ali “A Study of Technology in Firewall System” //IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA), Langkawi, Malaysia 2011, pp. 232-236

Fortinet: Global Leader of Cybersecurity Solutions and Services. https://www.fortinet.com/

Gerard Biau, Analysis of a Random Forests Model, Journal of Machine Learning Research 13 (2012) 1063-1095

Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel “Deep Learning for Anomaly Detection: A Review” 2021, pp.1-38

Hacirəhimova M.Ş.1, Yusifova L.R.2 “Experimental Study of Machine Learning Methods in Anomaly Detection” AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan, UOT 004.9:314/316

Hooshmand, M.K., Hosahalli, D.: Network anomaly detection using deep learning techniques. CAAI Trans. Intell. Technol. 7(2), 228–243 (2022).

https://blog.telegeography.com/total-international-bandwidth-now-stands-at-1217-tbps.

https://ml.cms.waikato.ac.nz/weka/.

Huang, Jin & Ling, Charles. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. Knowledge and Data Engineering, IEEE Transactions on. 17. 299- 310. 10.1109/TKDE.2005.50.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep learning. MIT press.

Imamverdiyev Y, Abdullayeva F (2018) Deep learning method for denial of service attack detection based on restricted Boltzmann machine. Big Data 6:2, 159–169, DOI: 10.1089/big.2018.0023.

İmamverdiyev Y.N "İnformasiya təhlükəsizliyi terminlərinin izahlı lüğəti"// Bakı: "İnformasiya Texnologiyaları nəşriyyatı" ,2015 ,160 səh.

Imamverdiyev, Y., & Abdullayeva, F. (2018). "Deep Learning Method for Denial of Service Attack Detection Based on Restricted Boltzmann Machine" //Big Data, 6(2), 159–169. doi:10.1089/big.2018.0023.

J. M. Bonifacio et al., "Neural Networks Applied in Intrusion Detection System," IEEE, 1998, pp. 205–10.

J. P. Anderson, "Computer Security Threat Monitoring and Surveillance," 1980.

Jiawei H., Micheline K. "Data Mining Concepts and Techniques" Third Edition Morgan Kaufmann Publishers is an imprint of Elsevier.

Jiong Z., Mohammad Z., and Anwar H. "Random-Forests-Based Network" //Intrusion Detection Systems. applications and reviews, 2008, vol. 38, no. 5, pp.649-659.

Joa o B.D. Cabrera, Carlos Gutie´rrez, Raman K. Mehra. "Ensemble methods for anomaly detection and distributed intrusion detection in Mobile" Scientific Systems Company, Inc., 500 West Cummings Park, Suite 3000, Woburn, MA 01801, United States.

Jurek A, Bi Y, Wu S, Nugent C. A survey of commonly used ensemble-based classification techniques. The Knowledge Engineering Review. 2014;29(5):551-581. doi:10.1017/S0269888913000155

K. J. Singh and T. De, "An approach of ddos attack detection using classifiers," Emerging Research in Computing, Information, Communication and Applications, 2015.

Kaufman, L., & Rousseeuw, P. J. (2005). "Finding groups in data: An introduction to cluster analysis" (Wiley series in probability and statistics). New York: Wiley-Interscience.Google Scholar.

Khaled Fawagreh, Mohamed Medhat Gaber & Eyad Elyan (2014) Random forests: from early developments to recent advancements, Systems Science & Control Engineering: An Open Access Journal, 2:1, 602-609, DOI: 10.1080/21642583.2014.956265

Krügel C, Toth T, Kirda E. Service specific anomaly detection for network intrusion detection. In: Proceedings of the 2002 ACM symposium on applied computing, SAC '02, ACM, New York, NY, USA; 2002. p. 201–208

Kumari, R., Sheetanshu, Singh, M. K., Jha, R., & Singh, N. K. ” Anomaly detection in network traffic using K-mean clustering” 2016 third International Conference on Recent Advances in Information Technology (RAIT). doi:10.1109/rait.2016.7507933

L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees. Chapman & Hall, New York, 1984.

L. Breiman. Bagging predictors. Machine Learning, 24:123–140, 1996.

L. Breiman. Consistency For a Simple Model of Random Forests. Technical Report 670, UC Berkeley, 2004.

L. Breiman. Random forests. Machine Learning, 45:5–32, 2001.

L. Breiman. Some Infinity Theory for Predictor Ensembles. Technical Report 577, UC Berkeley, 2000.

L. Feinstein, D. Schnackenberg, R. Balupari and D. Kindred, "Statistical approaches to DDoS attack detection and response," Proceedings DARPA Information Survivability Conference and Exposition, Washington, DC, USA, 2003, pp. 303-314 vol.1, doi: 10.1109/DISCEX.2003.1194894.

L. Portnoy, E. Eskin, S. Stolfo, Intrusion detection with unlabeled data using clustering, in: Proceedings of ACM CSS Workshop on Data Mining Applied to Security, DMSA-2001.

M. K. Islam, P. Hridi, M. S. Hossain and H. S. Narman, "Network Anomaly Detection Using LightGBM: A Gradient Boosting Classifier," 2020 30th International Telecommunication Networks and Applications Conference (ITNAC), Melbourne, VIC, Australia, 2020, pp. 1-7

M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

M.T. Dlamini, J.H.P. Eloff , M.M. Eloff “Information security: The moving target” //İnformation and Computer Security Architectures Research Group, Department of Computer Science, University of Pretoria, South Africa School of Computing, UNISA, Pretoria, South Africa //2008, pp.189-198

Maksutov, A. A., Cherepanov, I. A., & Alekseev, M. S. “Detection and prevention of DNS spoofing attacks”// 2017 Siberian Symposium on Data Science and Engineering (SSDSE), pp. 84-87

Manikopoulos, C., & Papavassiliou, S. (2002). Network intrusion and fault detection: a statistical anomaly approach. IEEE Communications Magazine, 40(10), 76–82. doi:10.1109/mcom.2002.1039860

Merrill Warkentin, Craig Orgeron “Using the security triad to assess blockchain technology in public sector applications” // International Journal of Information Management, 2020, pp.1-8

Michael E. Whitman and Herbert J. Mattord, “Principles of Information Security” // 2012, pp.1-617

Mina Eshak Magdy, Ahmed M. Matter, Saleh Hussin, Doaa Hassan, Shaimaa Ahmed Elsaid; “Anomaly-based intrusion detection system based on feature selection and majority voting”, Indonesian Journal of Electrical Engineering and Computer Science Vol. 30, No. 3, June 2023, pp. 1699~1706.

Mohammed Hussein Thwaini “Anomaly Detection in Network Traffic using Machine Learning for Early Threat Detection” 2022, pp.1-16

Mohiuddin Ahmed, Abdun Naser Mahmood, Jiankun Hu “A survey of network anomaly detection techniques”// Journal of Network and Computer Applications, 2015, pp. 1-13(anomaliya)

Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. “Network anomaly detection: methods, systems and tools”; IEEE Communications Surveys & Tutorials, 16(1):303–336, 2014

Münz G., Li S., Carle G. Traffic anomaly detection using k-means clustering. In: GI/ITG Workshop MMBnet. pp. 13{14 (2007)

Nauman Shahid, Ijaz Haider Naqvi, and Saad Bin Qaisar. “Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey. Artificial Intelligence Review”, 43(2):193–228, 2015.

Nikunj C. Oza a, Kagan Tumer, “Classifier ensembles: Select real-world applications”, NASA Ames Research Center, Mail Stop 269-2, Moffett Field, CA 94035-1000, United States.

Nina Godbole, SunitBelpure, Cyber Security Understanding Cyber Crimes, Computer Forensics and Legal Perspectives, Wiley // 2020 pp.1-38

Nong Ye, Qiang Chen; “An anomaly detection technique based on a Chi-square statistic for detecting intrusions into information systems”, Qual. Reliab. Engng. Int. 2001; 17: 105–112

P. Garcı´a-Teodoroa, J. Dı´az-Verdejoa, G. Macia´-Ferna´ndeza, E. Va´zquezb; “Anomaly-based network intrusion detection: Techniques, systems and challenges”

Payal Wadhwa, Sprinto: Continuous Security & Compliance Platform, 2024 https://sprinto.com/blog/types-of-security-models/

problem for classi□cation. Int. J. Comput. Appl. 127(15), 37–41 (2015)

Qasımov V.Ə. “İnformasiyanın qorunmasının müasir texnologiyaları” // MTN-in Heydər Əliyev adına Akademiyasının nəşriyyatı. 2011, 54-55

Ralph Foorthuis “On the nature and types of anomalies: a review of deviations in data”// International Journal of Data Science and Analytics 2021, pp. 297–331

Ramiz Alıguliyev, Makrufa Sh. Hajirahimova “Classification Ensemble Based Anomaly Detection in Network Traffic” // Review of Computer Engineering Research 2019, Volume 6, 1, pp 12-23.

S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian, “Cluster-ing without over-representation,” in Proceedings of the 25th ACMSIGKDD International Conference on Knowledge Discovery &DataMining, 2019, pp. 267–275.

S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, and F. Tang, “Discriminating ddos attacks from flash crowds using flow correlation coefficient,” IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 6, pp. 1073–1080, 2012.

Samreen Naeem1, Aqib Ali1, Sania Anam2and Muhammad Munawar Ahmed. “An Unsupervised Machine Learning Algorithms: Comprehensive Review”

Sarvesh Kumar, Upasana Gupta, Arvind Kumar Singh and Avadh Kishore Singh “Artificial Intelligence: Revolutionizing Cyber Security in the Digital Era”//Journal of Computers, Mechanical and Management, 2023, pp.31-42

Sergey Sakulin, Alexander Alfimtsev, Konstantin Kvitchenko, Leonid Dobkacz, Yuri Kalgin, Igor Lychkov; “Network Anomalies Detection Approach Based on Weighted Voting”, International Journal of Information Security and Privacy, Volume 16, Issue 1

several methods for balancing machine learning training data. ACM SIGKDD Explorations Newslett. 6(1), 20–29 (2004)

Shaik Akbar, Dr.K. Nageswara Rao and Dr.J.A. Chandulal “Intrusion Detection System Methodologies Based on Data Analysis”// International Journal of Computer Applications (0975 – 8887), Volume 5– No.2, August 2010, pp. 10-20

Sharmin Rashid, Subhra Prosun Paul “Proposed Methods of IP Spoofing Detection & Prevention” // 2013, pp

Sheenam, Abhinav Bhandari; “Chi-Square Statistical based Technique for Intrusion Detection”, International Journal of Security and Its Applications Vol. 10, No. 9 (2016), pp.87-98

Singh, A., Purohit, A.: A survey on methods for solving data imbalance

Smitha Rajagopal, Poornima Panduranga Kundapur, Katiganere Siddaramappa Hareesha, "A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets", Security and Communication Networks, vol. 2020, Article ID 4586875, 9 pages, 2020.

Srikanth Thudumu, Philip Branch, Jiong Jin and Jugdutt (Jack) Singh “A comprehensive survey of anomaly detection techniques for high dimensional big data” // Journal of Big data 2020, pp. 2-30

Sumeet Dua and Xian Du “Data Mining and Machine Learning in Cybersecurity”,2011, pp. 1-4

Surbhi Gupta, Abhishek Singhal and Akanksha Kapoor “A Literature Survey on Social Engineering Attacks: Phishing Attack”// International Conference on Computing, Communication and Automation (ICCCA) 2016, pp.537-540

T. Subbulakshmi, K. BalaKrishnan, S. M. Shalinie, D. AnandKumar, V. GanapathiSubramanian, and K. Kannathal, “Detection of ddos attacks using enhanced support vector machines with real time generated dataset,” in Third International Conference on Advanced Computing, pp. 17–22, 2011.

The CAIDA UCSD ”DDoS Attack 2007” Dataset, accessed by Jan 2018. http://www.caida.org/data/passive/ddos-20070804 dataset.xml.

Varun Chandola, Arindam Banerjee and Vipin Kumar “Anomaly Detection: A Survey”// ACM Computing Surveys, Vol. 41, No. 3, Article 15, Publication date: July 2009

Y.N. Imamverdiyev, L.V. Sukhostat, “Network trafic anomalies detection based on informative features”, Radio Electronics, Computer Science, Control, 2017, no.3, pp.113-119.

Yong Fang, Yunyun Zhang, Cheng Huang1, “Credit Card Fraud Detection Based on Machine Learning”, CMC, vol.61, no.1, pp.185-195, 2019.

Zhang J., Zulkernine M., and Haque A. Random-Forests-Based Network Intrusion Detection Systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2008, vol. 38, no. 5, pp. 649 – 659.