

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ

AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

YÜKSƏK TƏHSİL İNSTİTUTU

*Əlyazması hüququnda*

Məhərrəmov Aytac Müşfiq [Big data texnologiyaları ilə bank sektorunun strategiyalarının effektiv idarə edilməsi]

Yusifova Gülnar Asəf [Data mining metodlarının bank sektoruna tətbiqi]

Salmanova Aygül Həbil [Python proqramlaşdırma dilində kredit risk modelinin qurulması]

Şərifov Kemran Nofelyeviç [Python proqramlaşdırma dilində kredit risk modelinin qurulması]

BİG DATA TEXNOLOGİYALARINDA DATA MİNING METODLARININ  
TƏTBİQİ

mövzusunda

MAGİSTRİK DİSSERTASIYASI

İxtisas: 060509 – “Kompüter elmləri”

İxtisaslaşma: “İntellektual sistemlər”

Elmi rəhbər: t.ü.f.d, dosent Ağayev Firudin Tərhan

BAKİ – 2024

**AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ**  
**YÜKSƏK TƏHSİL İNSTİTUTU**

***MAGİSTRANTIN ANDI***

Big data texnologiyalarında data mining metodlarının tətbiqi mövzusunda təqdim etdiyim(iz)(Magistrlik dissertasiyasının mövzusu) magistrlik dissertasiyasını elmi əxlaq normalarına və istinad qaydalarına tam riayət etməklə və istifadə etdiyim bütün mənbələri ədəbiyyat siyahısında əks etdirməklə yazdığımıza and içirəm(ik) və magistrlik dissertasiyasının AzTU Kitabxana İnformasiya Mərkəzində saxlanması, həmin mərkəz tərəfindən AzTU Rəqəmsal Repozitoriyasına daxil edilərək repozitoriyanın veb saytında yerləşdirilməsinə icazə veririk.

**Məhərrəmov Aytac**

**Yusifova Gülnar**

**Salmanova Aygül**

**Şərifov Kemran**

**Tarix**

## MÜNDƏRİCAT

<b>GİRİŞ</b> .....	4
<b>I FƏSİL. BİG DATA TEXNOLOGİYALARI İLƏ BANK SEKTROUNUN STRATEGİYALARININ EFFEKTİV İDARƏ EDİLMƏSİ (MƏHƏRRƏMOVA AYTAC MÜŞFİQ)</b> .....	6
1.1. Bank sahəsində big data texnologiyaları.....	<b>Error! Bookmark not defined.</b>
1.2. Big data texnologiyalarının bank sektorunun idarə edilməsi və qərar qəbul edilməsində rolu .....	11
1.3. Big datada data mining inteqrasiyası .....	12
<b>II FƏSİL. DATA MİNING METODLARININ BANK SEKTORUNA TƏTBİQİ (YUSİFOVA GÜLNAR ASƏF)</b> .....	19
2.1. Məlumat mənbələri və toplanması .....	19
2.2. Bank sahəsində data mining üsulları və alətləri .....	22
2.3. Bank sektorunda risklərin idarə edilməsi .....	34
<b>III FƏSİL. PYTHON PROQRAMLAŞDIRMA DİLİNDƏ KREDİT RİSK MODELİNİN QURULMASI (SALMANOVA AYGÜL HƏBİL, ŞƏRİFOV KEMRAN NOFELYEVİÇ)</b> .....	50
3.1. Kredit risk modeli məlumatlarının təmizlənməsi və hazırlanması .....	50
3.2. Kredit risk modelinin inkişafı və qiymətləndirilməsi .....	59
<b>N Ə T İ C Ə</b> .....	69
<b>İSTİFADƏ EDİLMİŞ ƏDƏBİYYATLAR</b> .....	70
<b>S U M M A R Y</b> .....	73
<b>P E 3 İ O M E</b> .....	74

## İXTİSARLARIN SİYAHISI

AI – Artificial İntelligence

CRM - Customer Relationship Management

ETL - Extract , Transform, Load

FTK - Federal Ticarət Komissiyasının

IV – Information Value

İC – İnformasiya Cəmiyyəti

MAE – Mean Absolute Error

ML – Machine Learning

MSE – Mean Squared Error

PD - Probability of Default

WoE – Weight of Evidence

## GİRİŞ

**Mövzunun aktuallığı.** Big Data texnologiyaları günümüz iş dünyasında və digər sahələrdə böyük bir rəqabət üstünlüyü yaratmaq üçün çox mühüm bir rol oynayır. Big Data, böyük həcmli, qarışıq və müxtəlif data qruplarını emal etmə, depolama və təhlil etmə texnologiyalarını əhatə edərkən, bununla bərabər Data Mining isə bu data qruplarından lazımsız məlumatları çıxarmaq, gizli əlaqələri ortaya çıxarmaq və s üçün istifadə olunan bir sıra texnik və metodları ifadə edir. Data Mining, bir neçə texnikanın və alqoritmin kombinasiyasını tələb edən statistik analitika prosesidir ki, bu proses zamanı asan məlumatları istifadə edərək daha çətin, potensial olaraq lazımi məlumatlar aşkar edilir. Data Mining-in Big Data texnologiyaları ilə inetqrasiyası bugünkü gündə bizlərə bir çox üstünlük təklif edir:

1. Məlumat çıxarışı - Big Data daxilindəki böyük həcmli datanı effektiv şəkildə təhlil edərək, işlər və ya təşkilatlar üçün lazımi məlumatların çıxarılmasına imkan verir. Data Mining texnikləri, bu böyük data qrupundan lazımsız məlumatları açığa çıxara bilər, bu da bizlərə qərar qəbul proseslərini yaxşılaşdırmağa və strateji üstünlük əldə etməyə imkan verir.

2. Proqnoz və analitik bacarıqlar - Data Mining, böyük data qruplarında gələcəkdəki tendensiyaları proqnoz etmə bacarığına sahibdir. Məsələn, müştərinin davranışlarını anlamaq, marketinq strategiyalarını optimallaşdırmaq və ya əməliyyat effektivliyini artırmaq üçün istifadə oluna bilər.

3. Sənaye tətbiqləri - fərqli sənayelərdə tətbiq oluna bilər. Məsələn, maliyyə, pərakəndə, sağlamlıq, istehsal və s. kimi fərqli sahələrdə bu texnologiyaların tətbiqi, sənayələrə xüsusi ehtiyaclar üçün həllər təklif etməyə imkan verir.

4. Rəqabət üstünlüyü - təşkilatlara rəqabət üstünlüyü təmin edir. Düzgün təhlil və çıxarışlarla təşkilatlar daha yaxşı strateji qərarlar qəbul edə, müştəri məmnuniyyətini artırma və bazarda lider mövqe qazana bilərlər.

**Tədqiqatın məqsədi.** Geniş həcmli və çeşidlənməsi çətin olan məlumatlar arasında gizli əlaqələri və müşahidələri aşkar etməkdir. Bu metodlar vasitəsilə, mövcud məlumatlar analiz edilir və gizli trendlər, əlaqələr və nəzəriyyələr müəyyənləşdirilir.

Bu məlumatlar vasitəsilə fərqli sahələrdəki təşkilatların və ya layihələrin performansını artırmaq, strategiyalarını yaxşılaşdırmaq və qərarlarını dəstəkləmək mümkündür.

**Tədqiqatın predmeti və obyektı.** Bank sektoruna Data Mining tədqiqat sistemləri vasitəsilə məlumatların yaradılması prosesi onun keyfiyyətini təmin edən bir sıra standartlara cavab verməli olan vacib məsələlərdən biridir, çünki bank məlumatlarının tədqiqat sistemlərinin nəticələri ən yüksək nəticə əldə etmək üçün vacibdir. Bank məlumatları üçün tədqiqat sistemlərinin nəticələri ilə məşğul olan ən az vaxt və standartlara uyğun olmalıdır.

**Təcrübi əhəmiyyəti.** Banklarda kiber-risklərin idarə edilməsini təmin etmək üçün tədqiqat və məlumat axtarış sistemlərinin rolu, kiber-risklərin idarə edilməsində Data Mining alətlərinin tətbiqi təcrübi əhəmiyyətə malikdir. Banklarda mövcud olan bütün informasiya sistemləri arasında əlaqələrin gücləndirilməsi, kiber-risklərin yaxşı idarə edilməsi, müasir texnologiyala məxsus, lazımi metod və alətlərlə istifadə etmək olduqca zəruridir. Qeyd edilir ki, kommersiya banklarında məlumat axtarışı və axtarış sistemləri mühitin inkişafı ilə bilik sistemlərinin təkmilləşdirilməsində Data Mining verilənlərin emalı alətlərinin tətbiqi praktik əhəmiyyət kəsb edir.

**Elmi yeniliklər.** Data Mining bank sektoruna tətbiqi müxtəlif biznes qərarlarının qəbulu üçün əsas kimi xidmət edən nəticələr əldə etməyə imkan verir. Onların etibarlılığını yaxşılaşdırmaq və şirkətin mənfəətini artırmaq üçün geniş çeşidli məlumatlar toplanır və təhlil edilir. Oxşar müştərilər haqqında tarixi məlumatları emal etməklə şirkət riski qiymətləndirə və heç bir məlumatı olmayan potensial müştərilərin ömür boyu dəyərlərini proqnozlaşdırmağa bilər. Üstəlik, təsir edən amillər və nəticələnən göstəricilər həm aşkar, həm də gizli əlaqələrə malik ola bilər.

Bank sektorunda müştərinin biometrik identifikasiyası olması risklərin idarə edilməsində təhlükəsizliyi daha effektiv təmin edə bilər.

**Tədqiqat işinin strukturu və həcmi.** Dissertasiya işi girişdən, 4 fəsildən, nəticədən və 39 ədəbiyyat siyahısından ibarət olub, 74 səhifədə, 26 şəkildə verilib.

I fəsil bank sektorunda Big Data texnologiyalarının tətbiqi ilə bankların strateji qərarlarının effektiv idarə edilməsinin önəmini vurğulayır.

II fəsil Data Mining metodlarının bank sektorunda necə istifadə edildiyini və bankların məlumatları necə analiz etdiklərini araşdırır.

III fəsil bank sektorunda Big Data və Data Mining texnologiyalarının praktiki tətbiqi ilə bağlı təcrübələri və göstəriciləri tədqiq edir. Bu, tədqiqatın nəticələri ilə məlumatları birləşdirir və mövzunu daha aydın anlamağa kömək edir.

# I FƏSİL. BIG DATA TEXNOLOGİYALARI İLƏ BANK SEKTROUNUN STRATEGİYALARININ EFFEKTİV İDARƏ EDİLMƏSİ

## 1.1. Bank sahəsində big data texnologiyaları

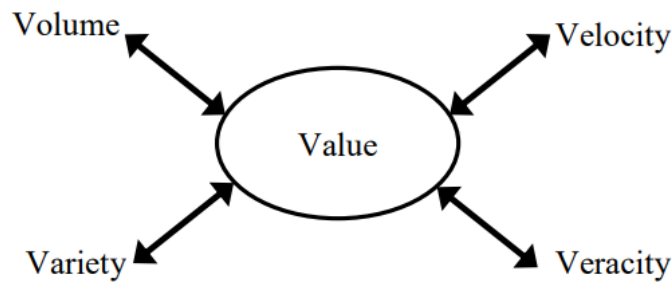
**Big Data.** İnformasiya, bilik və yüksək texnologiyalara əsaslanan informasiya cəmiyyəti (İC) böyük elmi-texniki inqilab çağında yer alır və bu inqilab, biznesdən başlayaraq elmi sahələrə qədər həyatımızın bütün sahələrinə nüfuz edir. Bu eyni zamanda sosial-iqtisadi, təhlükəsizlik və elmi məqamlar üçün mürəkkəb bir məsələdir. Problemi daha da dərinləşdirmək üçün "Big data" termini daxil olur. Bu yeni termin, mövcud idarəetmə metodları və ya intellektual analiz alətləri ilə əməl oluna bilməyən verilənləri identifikasiya etmək üçün istifadə olunur. Son illərdə artan verilənlərin saxlanması, idarə edilməsi və dəyər yaradılması ciddi bir problem yaradır. Məsələn, müxtəlif mənbələrdən gələn verilənlərin ənənəvi metodlar və alətlərlə əməl edilməməsi ilə bağlıdır. Bu səbəbdə də bu texnologiyaların elmi-tədqiqat obyektinə kimi öyrənilməsi vacibdir. Yüzlərlə terabayt və ekzabayt həcmində strukturlaşdırılmamış məlumatın toplanması, idarə edilməsi, saxlanması, təhlükəsizliyi, axtarışı, analizi (analitik hesabatların generasiyası və vizuallaşdırılması, proqnozlaşdırma) kimi məsələlərin həlli üçün yeni texnologiyalar, yanaşmalar və daha mükəmməl analiz üsulları tələb edir.

Big Data müxtəlif şirkətlər tərəfindən toplanan və emal edilə bilən verilən əldə etmək və ya Maşın Öyrənmə (Machine Learning) layihələri üçün istifadə edilə bilən böyük verilənlər dəstlərinə istinad edən bir termdir. Bu termin ilk dəfə 1998-ci ildə kompüter elmləri üzrə mütəxəssis olan Con Meşi tərəfindən istifadə edilmişdir. Lakin buna baxmayaraq, bu termin populyarlığını 2008-ci ildə həftəlik "Nature" jurnalının "big data" mövzusunda həsr olunmuş xüsusi nömrəsində Klifford Linçin (Clifford Lunch) "Big Data: How do your data grow?" adlı məqaləsi ilə qazanmışdır [Əliquliyev Rasim, Hacırahimova Məkrufə, 2014].

**Big Data xarakteristikaları.** Tətbiq olunduğu sahələrdən asılı olmayaraq, böyük verilənləri təsvir edən, həmçinin onların əsas problemlərini özündə əks etdirən bəzi ümumi xarakteristikalar mövcuddur: həcm (volume), sürət (velocity) və



müxtəliflik (variety). İngilis dilli mənbələrdə adətən bunu «3V»lər də adlandırırlar. Bu parametrlərin konvergensiyası böyük verilənləri təyin etməyə və digər verilənlərdən fərqləndirməkdə bizə kömək edir. Bu cür model ilk dəfə 2001-ci ildə D.Laney tərəfindən verilmişdir və “big data” texnologiyalarının əsas konsepsiyasını təşkil edir. Bu konsepsiya çox böyük sürətlə və müxtəlif mənbələrdən toplanan çox böyük həcmdə verilənləri daha səmərəli istifadə etmək, saxlamaq, analiz edərək ondan daha qiymətli informasiyanı əldə etmək ideyasını özündə əks etdirir. Bu üç tərif birləşdirməklə “5V” adlanan daha əhatəli tərif alınır: həcm(volume), sürət (velocity), müxtəliflik (variety), dəyər (value) , doğruluq (veracity).



Şək. 1.1. Big Data – 5V

Həcm – böyük verilənlərin əsas xarakteristikalarından biridir. Həcm probleminin əsasında saxlama problemləri yaranır, ki, bu da genişmiqyaslı saxlama və paylanmış emal tələb edir. Hazırda saxlama probleminin həllində informasiyanın qurğular arasında miqrasiyasını həyata keçirən bir sıra texnologiyalar mövcuddur: DAS (Direct-Attach Storage), NAS (Network Attached Storage), SAN (Storage Area Networks), HSM (Hierarchical Storage Management), ILM (Information Life-cycle Management). Həcmindən asılı olaraq böyük verilənlər üç qrupa bölünür: sürətli verilənlər (fast data) – onların həcmi terabaytlarla ölçülür; böyük analitika (big analytics) - onların həcmi petabaytlarla ölçülür; dərinə nüfuz etmə (deep insight) - onların həcmi ekzabaytlarla və zetabaytlarla ölçülür.

Sürət - həcm artmağa davam etdikcə, emal üçün də çox böyük sürət tələb olunur və burada iki hal nəzərdə tutulur. İlk növbədə, yeni verilənlər böyük sürətlə generasiya olunur, mövcud olanlar yenilənir və toplanır. İkincisi isə odur ki, sürət zaman problemi

kimi dəyərləndirilir və mövcud emal texnologiyalarının verilənləri real-vaxtda analiz etmək imkanına malik olması ilə izah olunur.

Müxtəliflik – böyük verilənlər, ənənəvi relyasiya verilənlər bazasının sətir və sütunlarında ifadə olunmuş strukturlaşdırılmış verilənlərlə birlikdə, strukturlaşdırılmamış tipli də ola bilər. Buna misal olaraq, mətn, video-audio fayllar, təsvirlər və s göstərmək olar. Bu növ verilənlər dünyada bütün informasiyanın 80-90%-ni təşkil edir. Bunları bir araya yığmaq və birgə emal etmək və analiz üçün onları uyğun şəkllə salmaq çox çətindir.

Doğruluq – böyük verilənlər öz öhdəliyini yerinə yetirməzdən əvvəl təmizlənməli və təşkil edilməli olan çoxlu sayda qeyri-müəyyən və çirkli, yəni yoxlanılmamış məlumatları özündə ehtiva edir. Məsələn, yarı strukturlaşdırılmış məlumatlar çox vaxt natamam və yaxud qeyri-dəqiq olur, bu da məlumatların təmizlənməsini çətinləşdirir. Qeyri-dəqiq və yaxud dublikat məlumatların toplanması ilə bağlı problemlər də var. Məlumat daxil edildikdə, çox zaman səhv və ya natamam transkripsiya edilir. Bu da yanlış nəticələrə gətirib çıxara bilər. Məlumatların keyfiyyətinin idarə edilməsi problemdir, çünki strukturlaşdırılmamış və ya yarı strukturlaşdırılmış məlumatlar tədqiqat layihəsinə bilik qazandıra bilməz. Bir verilənlər toplusunda bütün dəyişənlərə heç vaxt etibar etmək mümkün deyilsə, o zaman diqqəti yalnız ən etibarlı və ya etibarlı görünən dəyişənlərə yönəltmək vacibdir.

Dəyər – digər verilənlər böyük verilənlərə təsir edən xarici amilləri təmsil etdiyi halda, dəyər biznes strategiyası və icrası ilə bağlı daxili amilləri təmsil edir. Böyük verilənlərdən maksimum dəyər əldə etmək üçün şirkətlər və məlumat alimləri analizləri vasitəsilə nələrə nail olmaq istədikləri üçün aydın bir məqsədə sahib olmalıdırlar. Əgər bu müəyyən edilərsə, ondan sonra onlar hansı məlumatların toplanmalı olduğunu və ondan necə istifadə olunacağını müəyyən edə bilərlər [Ularu Elena, Puican Florina, Apostu Anca, Velicanu Manole, 2012].

Bank sektorunda böyük verilənlərin müsbət istiqamətdə olan gələcəyi danılmazdır və çoxsaylı imkanlarla ümidverici görünür. Texnologiya inkişaf etməyə davam etdikcə, həmçinin bankların böyük məlumat analitikasından necə istifadə edə

biləcəyi genişlənir, maliyyə institutları və onların müştəriləri üçün də daha aydın bir perspektiv yaradır.

Süni intellektə əsaslanan qərar qəbulu - süni intellekt və maşın öyrənmə alqoritmlərinin bizə dəstəklərindən biri odur ki, daha dəqiq və vaxtında qərarlar qəbul etmək üçün getdikcə daha çox böyük verilənlərlə işləyəcək. Kredit riskinin qiymətləndirilməsindən investisiya strategiyalarına, məlumatların keyfiyyətinin idarə edilməsinin optimallaşdırılmasına və risklərin idarə edilməsi proseslərinə qədər süni intellekt mürəkkəb proseslərin avtomatlaşdırılmasında mühüm rol oynayacaq, bununla da səmərəliliyi artıracaq və səhvləri azaldacaq.

Xüsusi müştəri təcrübəsi - böyük verilənlərin analitikası sayəsində gələcək daha fərdiləşdirilmiş və interaktiv müştəri təcrübəsi görəcək. Banklar müştəri ehtiyaclarını təxmin edə və müştəri məmnuniyyətini artıraraq daha yaxşı xidmətlər təklif edə bilər.

Real vaxt analitikası - texnologiya inkişaf etməyə davam etdikcə, real vaxt analitika imkanları təkmilləşəcək və banklara ani qərarlar qəbul etməyə imkan verəcək. Bu daha çox saxtakarlığın aşkarlanması kimi təcili tədbirlərin görülməsi tələb olunan sahələr üçün çox vacib olacaq.

Davamlılıq və sosial məsuliyyət - böyük verilənlər bankların daha sosial məsuliyyətli olmasına kömək etməkdə də mühüm rol oynayacaq. Analitika maliyyə institutlarına investisiyalarının ətraf mühitə və sosial təsirlərini anlamaqda kömək edə bilər və daha davamlı biznes təcrübələrinə səbəb ola bilər.

Bu deyilənlərdən aydın olur ki, böyük verilənlərin bank sektorunda transformasiya gücü danılmazdır. Müştəri təcrübələrində inqilab etməkdən tutmuş əməliyyat səmərəliliyinin artırılmasına və risklərin idarə edilməsinə qədər böyük verilənlər müasir bankçılıqda mümkün olanlar üçün yeni meyarlar müəyyən edir.

## **1.2. Big data texnologiyalarının bank sektorunun idarə edilməsi və qərar qəbul edilməsində rolu**

Big Data texnologiyaları bank sektorunda qərar qəbuletmə proseslərini xeyli təkmilləşdirə bilər. Böyük miqdarda strukturlaşdırılmış və strukturlaşdırılmamış

məlumatların toplanması və təhlili ilə banklar müştəri davranışı, maliyyə bazarları və əməliyyat səmərəliliyi haqqında dəyərli fikirlər əldə edə bilirlər.

Big Data texnologiyaları maşın öyrənməsi, proqnozlaşdırıcı modelləşdirmə və məlumatların əldə edilməsi də daxil olmaqla geniş alət və texnikaları əhatə edir. Bu texnologiyalar müştəri məlumatlarını təhlil etmək, tendensiyaları və nümunələri müəyyən etmək və marketinq, məhsulun inkişafı və risklərin idarə edilməsi ilə bağlı əsaslandırılmış qərarlar qəbul etmək üçün istifadə edilə bilər.

Bundan əlavə, fərdi marketinq kampaniyaları vasitəsilə banklar fərdi müştərilərin seçimləri, davranışları və ehtiyacları haqqında məlumat əldə etmək üçün böyük datadan istifadə edə bilirlər. Müştəri məlumatlarını təhlil edərək, banklar öz marketinq mesajlarını xüsusi müştəri segmentlərinə uyğunlaşdırma, cəlb etmə və konversiya nisbətlerini yaxşılaşdırma bilirlər. Məsələn, bank ev kreditləri ilə maraqlanan müştəriləri müəyyən etmək və onlara müvafiq məlumat və təkliflərlə hədəflənmiş marketinq mesajları göndərmək üçün big datadan istifadə edə bilər. Bu, öz növbəsində, banklara daha məqsədyönlü və uyğun məhsul və xidmətlər təklif etməyə kömək edir, nəticədə müştəri məmnuniyyəti və loyallığı yüksəlir.

Müştərilərin segmentasiyası bank sektorunda böyük məlumatların digər əsas faydasıdır. Müştəriləri yaş, gəlir və əməliyyat tarixçəsi kimi müxtəlif faktorlara əsasən müxtəlif qruplara bölməklə banklar öz müştərilərinin ehtiyaclarını və seçimlərini daha yaxşı anlaya bilirlər. Xülasə, Big Data texnologiyaları müştəri davranışı, maliyyə bazarları və əməliyyat səmərəliliyi haqqında dəyərli fikirlər təqdim etməklə bank sektorunda qərar qəbul etmə proseslərini xeyli təkmilləşdirə bilər. Bu texnologiyalardan istifadə etməklə banklar saxtakarlığın aşkarlanmasını təkmilləşdirə, riskləri idarə edə, müştəri xidmətlərini yaxşılaşdırma və bazarda rəqabət üstünlüyü əldə edə bilər. Bank sektorunda istifadə oluna bilən bəzi xüsusi Big Data texnologiyaları bunlardır:

Python məlumatların təhlili və maşın öyrənməsi üçün məşhur proqramlaşdırma dilidir və böyük məlumatların emalı üçün faydalı olan çoxlu kitabxana və freymvorklərə malikdir. Məsələn, Pandas verilənlərin manipulyasiyası və təhlili üçün güclü kitabxanadır, Scikit-learn isə təsnifat, reqressiya, klasterləşdirmə və ölçülərin

azaldılması üçün müxtəlif alqoritmlər təqdim edən maşın öyrənmə kitabxanasıdır. Bundan əlavə, PySpark, Python istifadəçilərinə geniş miqyaslı məlumatların işlənməsi üçün Spark-ın gücündən istifadə etməyə imkan verən Spark üçün Python API-dir.

Digər tərəfdən, Oracle SQL strukturlaşdırılmış məlumatların sorğulanması və manipulyasiyası üçün güclü vasitədir. Böyük verilənlərin emalı üçün müxtəlif mənbələrdən məlumatları çıxarmaq və çevirmək üçün istifadə edilə bilər. Oracle həmçinin, istifadəçilərə proqramlaşdırma bacarıqları tələb etmədən Big Datanı tədqiq etməyə, təhlil etməyə və vizuallaşdırmağa imkan verən veb əsaslı vizual analitika vasitəsi olan Big Data Discovery alətini təqdim edir.

Python və Oracle SQL ilə yanaşı, bank sektorunda istifadə edilə bilən digər Big Data texnologiyalarına Hadoop, Apache Spark və MongoDB və Cassandra kimi NoSQL verilənlər bazaları daxildir. Bu texnologiyalar geniş miqyaslı məlumatların işlənməsi və təhlili üçün paylanmış saxlama və emal imkanlarını təmin edir.

Big Data texnologiyaları bank sektorunda qərar qəbul etmək üçün çoxsaylı imkanlar təqdim etsə də, bir sıra çətinliklər və mülahizələr də gətirir. Banklar məlumatların idarə edilməsi siyasətlərini yaratmalı, məlumatların keyfiyyətini və düzgünlüyünü təmin etməli, məlumatların məxfiliyi qaydalarına əməl etməli, müxtəlif mənbələrdən məlumatları inteqrasiya etməli, böyük həcmdə məlumatı idarə etməli, normativ tələblərə əməl etməlidir [Jagadish Hosagrahar, 2015].

### 1.3. Big datada data mining inteqrasiyası

**Data Mining.** Müasir rəqəmsal əsrdə hər gün milyardlarla məlumat yaradılır. İnternet axtarışları, sosial media paylaşımaları, satınalmalar və daha çox məlumat daim böyüyən okeana töhfə verir. Beləliklə, bu böyük məlumat yığımında qiymətli məlumatları necə kəşf edə bilərik? Burada Data Mining işə düşür.

Data Mining böyük verilən dəstlərindən lazımi məlumat çıxarmaq üçün statistik və riyazi üsulların istifadə edildiyi bir intizamdır. Data Mining bu nəhəng məlumat bazasında nümunələri, əlaqələri və meylləri kəşf etməyə kömək edir. Bu prosesdə

maşın öyrənməsi, statistik analiz, verilənlər bazası idarə edilməsi və məlumatların vizuallaşdırılması kimi üsullardan istifadə edilir.

Data Mining bir çox müxtəlif sənaye sahələrində istifadə olunur. Data Mining metodları səhiyyə, maliyyə, pərakəndə satış, telekommunikasiya və marketing kimi sahələrdə böyük uğur qazanmışdır. Məsələn, bank saxtakarlığın aşkarlanması üçün məlumatların öyrənilməsi üsullarından istifadə edə bilər və ya xəstəxana xəstəlik diaqnozlarını yaxşılaşdırmaq üçün böyük məlumat analitikasından istifadə edə bilər.

Məlumatların çıxarılması prosesi adətən aşağıdakı addımlardan ibarətdir:

- Məlumatların toplanması - məlumatlar müvafiq məlumat mənbələrindən toplanır və məlumat toplusu yaradılır.
- Məlumatların əvvəlcədən işlənməsi - məlumat dəsti təmizlənir, çatışmayan məlumatlar doldurulur və lazımsız məlumatlar silinir.
- Modelləşdirmə - verilənlər dəsti üçün uyğun model seçilir və bu modelə məlumatlar tətbiq edilir.
- Qiymətləndirmə - modelin dəqiqliyi və effektivliyi qiymətləndirilir.
- Nəticələrin şərhə - əldə edilmiş nəticələr təhlil edilir və lazımi məlumatlar çıxarılır.

Data Mining-də verilənlərin inteqrasiyası birdən çox mənbədən alınan məlumatların vahid görünüşdə birləşdirilməsi prosesinə aiddir. Bu, məlumatların təmizlənməsi və dəyişdirilməsi, həmçinin müxtəlif mənbələr arasında mövcud ola biləcək hər hansı uyğunsuzluq və ya münaqişələrin həllini əhatə edə bilər. Məlumat inteqrasiyasının məqsədi təhlil və qərar qəbulu məqsədləri üçün məlumatları daha faydalı və mənalı etməkdir. Verilənlərin inteqrasiyasında istifadə olunan üsullara məlumatların saxlanması, ETL (Extract - çıxarma, Transform - çevirmə, Load - yükləmə) prosesləri və məlumatların birləşməsi daxildir [Tan Pang, Steinbach Michael, Kumar Vipin, 2021].

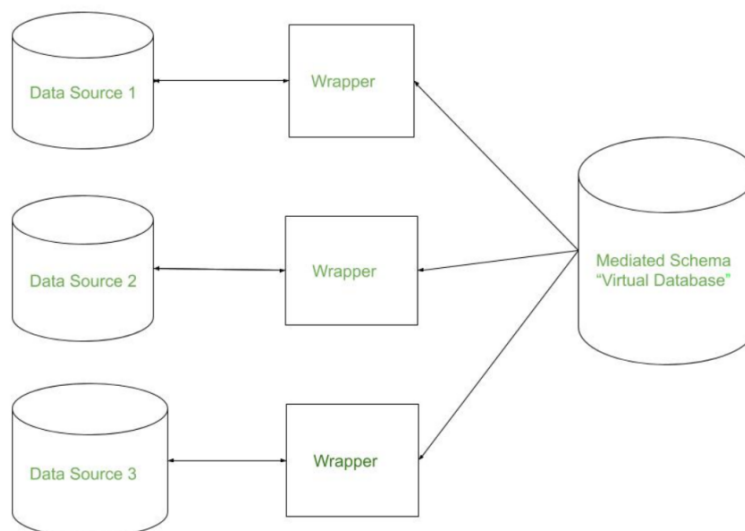
Verilənlərin inteqrasiyası çoxlu fərqli məlumat mənbələrindən məlumatları ardıcıl məlumat anbarında birləşdirən və verilənlərin vahid görünüşünü təmin edən verilənlərin əvvəlcədən işlənməsi texnikasıdır. Bu mənbələrə çoxsaylı məlumat topluları, verilənlər bazası və s daxil ola bilər.

Verilənlərin inteqrasiyası yanaşmaları formal olaraq üçlü  $\langle G, S, M \rangle$  kimi müəyyən edilir, burada,

- G qlobal sxem üçün dayanır
- S sxemin heterojen mənbəyini ifadə edir
- M mənbə sorğuları və qlobal sxem arasında xəritələşdirmə deməkdir

**Məlumat inteqrasiyası.** Məlumat inteqrasiyası bir çox mənbədən alınan məlumatların vahid və ardıcıl görünüşdə birləşdirilməsi prosesidir. Bu proses müxtəlif məlumat mənbələrinin müəyyən edilməsini və əldə edilməsini, verilənlərin ümumi formata uyğunlaşdırılmasını və mənbələr arasında hər hansı uyğunsuzluğun uzlaşdırılmasını əhatə edir. Məlumat inteqrasiyasının məqsədi verilənlərin daha tam və dəqiq başa düşülməsi üçün çoxsaylı sistemlər və ya platformalar arasında yayılmış məlumatlara daxil olmağı və təhlil etməyi asanlaşdırmaqdır.

Məlumat inteqrasiyası biznes kəşfiyyatı, məlumatların saxlanması, əsas məlumatların idarə edilməsi və analitika kimi geniş tətbiqlərdə istifadə olunur. Məlumat inteqrasiyası bu tətbiqlərin uğuru üçün kritik ola bilər, çünki o, təşkilatlara daha yaxşı qərarlar qəbul etmək, əməliyyat səmərəliliyini artırmaq və rəqabət qabiliyyətli olmaq üçün müxtəlif sistemlər, departamentlər və iş sahələrində yayılmış məlumatlara daxil olmaq və təhlil etmək imkanı verir.



Şək. 1.2. Məlumat inteqrasiyası sxemi

Məlumatların inteqrasiyası üçün iki əsas yanaşma mövcuddur – biri “sıx birləşmə yanaşması”, digəri isə “boş birləşmə yanaşması”dır.

- Sıx birləşmə (Tight coupling approach) - bu yanaşma inteqrasiya olunmuş məlumatları saxlamaq üçün mərkəzləşdirilmiş repozitoriya və ya məlumat anbarının yaradılmasını nəzərdə tutur. Məlumatlar müxtəlif mənbələrdən çıxarılır, çevrilir və məlumat anbarına yüklənir. Məlumatlar bir-biri ilə sıx əlaqəli şəkildə inteqrasiya olunur, yəni verilənlər bütün verilənlər bazası və ya sxem səviyyəsində olduğu kimi yüksək səviyyədə inteqrasiya olunur. Bu yanaşma həm də məlumat anbarı kimi tanınır və o, məlumatların ardıcılığını və bütövlüyünü təmin edir, lakin o, dəyişməz, dəyişdirilməsi və ya yenilənməsi çətin ola bilər. Burada məlumat anbarına məlumat axtarış komponenti kimi baxılır. Bu birləşmədə məlumatlar ETL prosesi vasitəsilə müxtəlif mənbələrdən vahid fiziki məkanda birləşdirilir.

- Boş birləşmə (Loose coupling approach) - bu yanaşma fərdi məlumat elementləri və ya qeydlər səviyyəsində ən aşağı səviyyədə məlumatların inteqrasiyasını nəzərdə tutur. Məlumatlar sərbəst şəkildə birləşdirilir, yəni verilənlər aşağı səviyyədə inteqrasiya olunur və bu, mərkəzi repozitoriya və ya məlumat anbarı yaratmadan məlumatların inteqrasiyasına imkan verir. Bu yanaşma həm də məlumat birləşməsi kimi tanınır və o, məlumat çevikliyi və asan yeniləmələri təmin edir, lakin çoxsaylı məlumat mənbələrində ardıcılığı və bütövlüyü qorumaq çətin ola bilər.

Burada istifadəçidən sorğu alan, onu mənbə verilənlər bazasının başa düşəcəyi şəkildə çevirən, sonra isə nəticəni əldə etmək üçün sorğunu birbaşa mənbə verilənlər bazalarına göndərən interfeys təmin edilir və məlumatlar yalnız faktiki mənbə verilənlər bazalarında qalır [Han Jiawei, Kamber Micheline, Pei Jian, 2021].

**Məlumat inteqrasiyasında problemlər.** Bir neçə mənbədən məlumatları birləşdirərkən bəzi problemlər yarana bilər, o cümlədən:

- Data keyfiyyəti (Data quality) - məlumatlardakı uyğunsuzluqlar və səhvlər birləşməni və təhlili çətinləşdirə bilər.

- Verilənlərin semantikasi (Data semantics) - fərqli mənbələr eyni məlumat üçün fərqli termin və ya təriflərdən istifadə edə bilər ki, bu da məlumatları birləşdirməyi və başa düşməyi çətinləşdirir.



- Verilənlərin heterojenliyi (Data heterogeneity) - fərqli mənbələr müxtəlif məlumat formatlarından, strukturlarından və ya sxemlərindən istifadə edə bilər ki, bu da məlumatları birləşdirməyi və təhlil etməyi çətinləşdirir.

- Məlumatların məxfiliyi və təhlükəsizliyi (Data privacy and security) - məxfi məlumatların qorunması və təhlükəsizliyin təmin edilməsi bir çox mənbədən verilənlərin inteqrasiyası zamanı çətin ola bilər.

- Müəyyən şərt daxilində ölçünün dəyişdirilməsi (scalability) - çoxlu mənbələrdən böyük həcmdə məlumatların inteqrasiyası hesablama baxımından bahalı və vaxt aparan ola bilər.

- Verilənlərin idarə edilməsi (Data governance) - bir çox mənbədən məlumatların inteqrasiyasını idarə etmək və saxlamaq çətin ola bilər, xüsusən də məlumatların dəqiqliyi, ardıcılığı və vaxtında təmin edilməsinə gəldikdə.

- Performans (Performance) - çoxsaylı mənbələrdən məlumatların inteqrasiyası da sistemin işinə təsir edə bilər.

- Mövcud sistemlərlə inteqrasiya (Integration with existing systems) - yeni məlumat mənbələrinin mövcud sistemlərlə inteqrasiyası əhəmiyyətli səy və resurslar tələb edən mürəkkəb bir iş ola bilər.

- Mürəkkəblilik (Complexity) - çoxsaylı mənbələrdən məlumatların inteqrasiyasının mürəkkəbliyi yüksək ola bilər, xüsusi bacarıq və bilik tələb edir.

Verilənlərin inteqrasiyası zamanı nəzərə alınmalı üç məsələ var: sxem inteqrasiyası, problemlərin aşkarlanması və məlumat dəyəri ziddiyyətlərinin həlli. Bunlar aşağıda qısa şəkildə izah olunur.

1. Sxem inteqrasiyası - müxtəlif mənbələrdən metadata birləşdirilməlidir. Çoxsaylı mənbələrdən olan real dünya obyektləri obyektin identifikasiyası problemi adlanır.
2. Problemlərin aşkarlanması - atribut başqa atributdan və ya atributlar dəstindən əldə edilə və ya əldə edilə bilsə, lazımsız ola bilər. Atributlardakı uyğunsuzluqlar da nəticədə əldə edilən məlumat dəstində problemə səbəb ola bilər. Bəzi problemlər korrelyasiya təhlili ilə aşkar edilə bilər.

3. Məlumat dəyəri ziddiyyətlərinin həlli - bu, məlumat inteqrasiyasında üçüncü kritik məsələdir. Fərqli mənbələrdən atribut dəyərləri eyni real dünya obyektini üçün fərqli ola bilər. Bir sistemdəki atribut digər sistemdəki "eyni" atributdan daha aşağı səviyyədə qeyd oluna bilər.

## II FƏSİL. DATA MİNING METODLARININ BANK SEKTROUNA TƏTBİQİ

### 2.1. Məlumat mənbələri və toplanması

Bank sektorunda məlumat mənbələri çoxsaylıdır və bunlar müştərilərdən, müxtəlif bank əməliyyatlarından, kredit məlumatlarından, rəqəmsal kanallardan, bazar məlumatlarından, xarici mənbələrdən və bankların öz risk idarəetmə sistemlərindən gələ bilər. Bu məlumatlar, bankların operativ fəaliyyətlərini idarə etmələri, riskləri qiymətləndirmələri, müştəri təcrübəsini yaxşılaşdırmaq üçün strateji qərarlar qəbul etmələri və müştəri tələblərinə cavab verərək məhsul və xidmətlərini təkmilləşdirmələri üçün çox mühüm rol oynayır. Bu deyilənləri daha da genişləndirsək, bəzi əsas məlumat mənbələri aşağıdakılardır:

- Müştəri əməliyyatları - müştəri əməliyyatlarından cəmlənən məlumatlar, çıxarışlar, depozitlər, köçürmələr və alışverişlər kimi, müştəri davranışları və istəkləri barədə anlayışlar verir.

- Hesab məlumatları - müştəri hesabları ilə əlaqəli məlumatlar, hesab növləri, əməliyyat tarixləri və hesab açılışları (bağlanmaları), balanslar kimi, müştəri münasibətlərinin və maliyyə əməliyyatlarının anlaşılması üçün əhəmiyyətlidir.

- Kredit məlumatları - banklar potensial borcalanların kreditə uyğunluğunu qiymətləndirmək üçün adiyatla kredit hesabatlarını əldə edirlər. Bu hesabatlar, fərdlərin kredit tarixləri, qalıq borcları və kredit reytingləri kimi məlumatlarla bağlı məlumatlar daxil edir.

- Bazar məlumatları - bazar məlumatları, faiz dərəcələri, valyuta məzənnələri və iqtisadi göstəricilər daxil olmaqla, bankların investisiyalar, ticarət fəaliyyətləri və risk idarəetməsi haqqında lazımi qərarlar vermələrinə kömək edir.

- Risk idarəetmə sistemləri - banklar, kredit riski, likvidlik riski, bazar riski və əməliyyat riski kimi müxtəlif növlərdə riskləri izləyən və analiz edən risk idarəetmə sistemləri vasitəsilə məlumatları daxil edirlər.

- Müştəri əlaqəsi və sorğuları - müştəri sorğuları, şikayətlər və müştəri xidməti nümayəndələri ilə qarşılıqlı əlaqədən əldə edilən rəy müştəri məmnuniyyəti, seçimlər və təkmilləşdirilməli sahələr haqqında çox önəmli və lazımi fikirlər təmin edir.

- Xarici məlumat təchizatçıları - banklar həmçinin bazar tendensiyaları, tənzimləmə tələbləri və müştəri demoqrafikası haqqında anlayışlarını zənginləşdirmək üçün dövlət qurumları, tənzimləyici orqanlar, üçüncü tərəf təchizatçıları və sosial media platformaları kimi xarici mənbələrdən məlumatlardan istifadə edə bilirlər.

- Rəqəmsal kanallar - onlayn bankçılıq platformaları, mobil proqramlar kimi rəqəmsal bankçılıq kanallarından əldə edilən məlumatlar rəqəmsal məkanda müştəri davranışı, seçimləri və s haqqında fikirlər təklif edir.

Bu cür müxtəlif mənbələrdən lazımi məlumatları effektiv şəkildə istifadə edərək, banklar müştərilərini daha yaxşı anlamaq, riskləri effektiv şəkildə idarə etmək və strateji qərarlarını qəbul etmək üçün lazımi qərarlar verməyə imkan verir.

Bank sahəsində məlumat mənbələri və məlumatların toplanmasında mühüm rola malik olan texnologiyalardan biri də Oracle-dır. Oracle, əlaqəli verilənlər bazası idarəetmə sistemi (RDBMS) olan Oracle Database daxil olmaqla, geniş çeşidli biznes yönümlü məhsul və xidmətləri təklif edən ABŞ-da yerləşən informasiya texnologiyaları şirkətidir. Banklar tez-tez müxtəlif mənbələrdən məlumatları toplayır və onları Oracle-ın məlumatların idarə edilməsi həlləri ilə inteqrasiya edirlər. Oracle banklara məlumatları təhlükəsiz saxlamağa, təhlil etməyə və məlumatlara əsaslanan qərarlar qəbul etməkdə kömək edir. Oracle-ın bank sektorunda məlumat mənbələrinin toplanmasında rol oynadığı bəzi vacib sahələr bunlardır:

- Müştəri məlumatları - Oracle banklara müştəri məlumatlarını toplamaqda və saxlamaqda kömək edir. Bu, müştəri əlaqələrini idarə etmək, fərdiləşdirilmiş xidmətlər göstərmək və müştəri təcrübəsini təkmilləşdirmək üçün çox vacibdir.

- Maliyyə məlumatları - banklar maliyyə bazarlarından və onların daxili əməliyyatlarından məlumat toplayır. Oracle həlləri bu məlumatları təhlil etmək və maliyyə fəaliyyətini qiymətləndirmək üçün istifadə olunur.

- Risk Məlumatları - banklar kredit riski, əməliyyat riski, bazar riski və digər risk növləri haqqında məlumat toplayır. Bu məlumatları izləmək və təhlil etmək və riskləri idarə etmək üçün Oracle-ın risklərin idarə edilməsi həllərindən istifadə olunur.

Oracle-ın bu məlumat mənbələrini toplamaq və idarə etmək imkanları banklara məlumatlara əsaslanan qərarlar qəbul etməyə və əməliyyat səmərəliliyini artırmağa çox böyük təkan verir. Buna görə də Oracle bank sektorunda mühüm bir rola malikdir.

Oracle, bankların müştəri məlumatlarını, əməliyyat tarixçələrini, risk məlumatlarını, maliyyə məlumatlarını və digər məlumatları toplamaq, saxlamaq və idarə etmək üçün bizə bir çox məhsul və xidmətlər təklif edir. Bunlardan bir neçəsinə nəzər salaq:

- Oracle Database - bankların ən kritik məlumatlarını saxlamaq və idarə etmək üçün istifadə olunan mövzuda çox öndə gedən bir məlumat bazası sistemidir. Banklar, müştəri hesab məlumatlarını, tranzaksiya tarixçələrini, kredit məlumatlarını və digər əhəmiyyətli məlumatları Oracle Database vasitəsilə saxlayır və əlçatan edirlər.
- Oracle Exadata - yüksək performans və məlumat bazası sistemi optimizasiyası ilə məşhurdur. Banklar, məlumat toplamaq üçün Exadata avadanlığını istifadə edərək, həm də məlumatları sürətli və effektiv şəkildə işləmək üçün yüksək performans və təhlükəsizlik təmin edə bilər.
- Oracle Data Integrator (ODI) - bankların fərqli mənbələrdən məlumatları toplamaq, birləşdirmək və təmizləmək üçün istifadə olunan bir məlumat integrasiya sistemidir. Bu, bankların müştəri məlumatlarını, bazar məlumatlarını və digər məlumatları bir araya gətirməsinə kömək edir.
- Oracle Big Data Appliance - bankların geniş həcmli və müxtəlif məlumatlardan istifadə etmələrini təmin edən və məlumat analitikasını inkişaf etdirmək üçün istifadə olunan bir platformdur. Banklar, bu platform vasitəsilə böyük məlumat kütlələrindən istifadə edərək, riskləri qiymətləndirə, müştəri təcrübəsini yaxşılaşdırı və iş effektivliyini artırabilir.
- Oracle Cloud - Oracle, bankların məlumatları saxlamaq və idarə etmək üçün bizə bir sıra bulud xidmətləri də təklif edir. Bu, bankların məlumatlara istənilən vaxtda, istənilən yerdə əlçatan olmalarına kömək edir və onlara məlumatların ehtiyat nüsxəsi və təhlükəsizliyi ilə bağlı problemləri həll etmək imkanı verir.

Oracle Database ilə bank sektorunda məlumat toplama əməliyyatlarını yerinə yetirmək istəsək, biz aşağıdakı addımları yerinə yetirməliyik:

1. Oracle verilənlər bazasının quraşdırılması – ilk öncə Oracle verilənlər bazası quraşdırılmalı olacağıq. Buraya Oracle Database proqramını yükləmək, quraşdırmaq və konfigurasiya etmək daxildir. Oracle Database məlumatların saxlanması və idarə olunması üçün əsas infrastrukturunu təmin edir.
2. Məlumatların modelləşdirilməsi və cədvəllərin yaradılması - bank məlumatlarını saxlamaq üçün lazımi cədvəllər yaradılmalıdır. Məsələn, müştəri məlumatlarını saxlamaq üçün "Customer" cədvəli, hesab əməliyyatlarını izləmək üçün "Account\_Transactions" cədvəli və s.
3. Məlumatların yüklənməsi və inteqrasiyası - Oracle-a müxtəlif mənbələrdən verilənləri yükləmək üçün Oracle tərəfindən təqdim olunan alətlərdən istifadə edə bilərik. Məsələn, SQL\*Loader və ya Oracle Data Pump kimi alətlər böyük verilənlərin yüklənməsi üçün istifadə edilə bilər. Biz həmçinin Oracle Integration Cloud kimi bulud əsaslı inteqrasiya platformaları ilə müxtəlif sistemlərdən məlumatları idxal edə bilərik.
4. Məlumat analizi və iş intellekti – Oracle-ın iş intellekti və analitik alətləri ilə məlumatları analiz edə bilər və lazımi məlumatlar əldə edə bilərik. Oracle Business Intelligence Enterprise Edition (OBIEE) kimi alətlər, interaktiv hesabatlar və vizuallaşdırma yaratmağa imkan verir.
5. Tənzimləyici uyğunluq və hesabatçılıq - bank sektorunda tənzimləyici tələblərə uyğunluq üçün Oracle-ın tənzimləyici uyğunluq alətlərindən istifadə edilə bilər. Bu alətlər, KYC, AML kimi tənzimləmələrə uyğunluğu təmin etməyi və tənzimləyici orqanlara hesabat verməyi asanlaşdırır.
6. Təhlükəsizlik və ehtiyat nüsxəsi - Oracle verilənlər bazasının təhlükəsizliyi üçün lazımi tədbirləri görülməlidir. Buraya giriş nəzarətləri, məlumatların şifrələnməsi kimi tədbirlər daxildir. Biz həmçinin müntəzəm məlumat ehtiyat nüsxəsini çıxarmaqla məlumat itkisinin qarşısını ala bilərik.

Bank sektorunda məlumatların toplanmasının uğurla başa çatması üçün, bank məlumatları effektiv şəkildə toplanmalı və Oracle Database-də təşkil edilməlidir. Müştəri məlumatları, hesab əməliyyatları və digər vacib məlumatlar təhlükəsiz şəkildə idarə oluna bilən edilməlidir [Madden Samuel, 2012].

## 2.2. Bank sahəsində data mining üsulları və alətləri

Data Mining, dəqiq qərarlar qəbul etmək üçün istifadə edilə biləcəyimiz nümunələr və əlaqələr kimi böyük məlumat dəstlərində məlumatların kəşf edilməsi prosesi və ya texnikasıdır. Bu proses kompüterlərin və süni intellekt və maşın öyrənməsi sayəsində mümkün olan avtomatlaşdırılmış proseslərin köməyi ilə baş verir. Data Mining, məqsədin nə olduğu sualından başlayaraq və strategiyaların işlənilib hazırlanmasına töhfə verən anlayışlarla bitən xüsusi bir prosesi izləyir.

Data Mining vacib xüsusiyyətlərindən biri odur ki, Data Mining vasitəsilə məlumatlarda tendensiyaları müəyyən etdikdən sonra biz daha çox müştəri əldə etmək, biznes əməliyyatlarını daha səmərəli etmək və istehlakçıları daha yaxşı başa düşmək üçün bu məlumatdan istifadə etməkdir. Hər bir sektor Data Mining-dən faydalana bilər, çünki onun tibbi diaqnostika, qiymət optimallaşdırılması, risklərin qiymətləndirilməsi və saxtakarlığın aşkarlanması və s. kimi tətbiqləri var.

Məlumatların öyrənilməsi üsulları müxtəlifdir və geniş şəkildə üç qrupa təsnif edilə bilər: nəzarət olunan, nəzarətsiz və yarı nəzarət olunan . Hər bir qrupun xüsusi məlumat təhlili növləri üçün uyğunlaşdırılmış fərqli metodları var.

- Nəzarət olunan texnikalar - nəzarət edilən öyrənmə, nəticə dəyişəninin (və ya "etiket") məlum olduğu etiketli verilənlər bazasında modeli öyrətməyi əhatə edir. Daha sonra model yeni, görünməmiş məlumatlar əsasında proqnozlar verir. Təsnifat, yəni bu üsul keçmiş məlumatlara əsaslanaraq yeni məlumatların kateqoriyasını və ya sinfini proqnozlaşdırır. Məsələn, banklar kredit tarixçəsinə əsasən kredit ərizəçisinin “yüksək riskli” və ya “az riskli” kateqoriyasına aid olduğunu müəyyən etmək üçün təsnifatdan istifadə edilə bilər. Reqressiya, yəni reqressiya üsulları bir və ya daha çox proqnozlaşdırıcı dəyişəne

əsaslanan davamlı nəticə dəyişənini proqnozlaşdırır. Məsələn, bank müştərinin keçmiş maliyyə davranışlarına əsaslanaraq gələcək kredit xalını proqnozlaşdırmaq üçün reqressiyadan istifadə edə bilər.

- Nəzarətsiz texnikalar - nəzarətsiz öyrənmə, nəticə dəyişəninin məlum olmadığı verilənlər bazası ilə məşğul olur. Burada məqsəd verilənlərin strukturunu araşdırmaq və əvvəlcədən təyin edilmiş etiketlər olmadan nümunələri və ya qrupları tapmaqdır. Klasterləşdirmə, yəni xas oxşarlıqları paylaşan məlumat nöqtələrini qruplaşdırmaqdır. Bank müştəriləri xərcləmə vərdişlərinə görə müxtəlif qruplara ayıra bilər ki, bu da daha məqsədyönlü marketing kampaniyalarını asanlaşdırır. Nümunələrin tanınması, yəni bu texnika verilənlərdəki nümunələri və qanunauyğunluqları müəyyən edir. O, məsələn, maliyyə əməliyyatlarında təkrarlanan tendensiyaları tanımaq üçün istifadə edilə bilər. Anomaliya və outliers aşkarlanması, yəni anomaliya aşkarlanması normadan əhəmiyyətli dərəcədə kənara çıxan məlumat nöqtələrini müəyyən edir. Bank işində bu, saxta əməliyyatları aşkar etməyə kömək edə bilər. Bu, ilk növbədə nəzarət olunmayan bir üsul olsa da, etiketli məlumatımız varsa, anomaliyaların aşkarlanmasına nəzarət edilən öyrənmə vasitəsilə də yanaşmaq olar.
- Yarı nəzarətli texnikalar - yarı nəzarətli öyrənmə nəzarət edilən və nəzarətsiz öyrənmə arasında olur. O, təlim üçün həm etiketli, həm də etiketsiz məlumatlardan istifadə edir ki, bu da tam etiketlenmiş verilənlər toplusunun alınması bahalı və ya vaxt aparan zaman onu xüsusilə faydalı edir. Bank işi kontekstində, etiketlenmiş əməliyyat məlumatlarının (saxtakar və ya qanuni) və daha çox etiketsiz əməliyyatlar toplusunun qarışığı olduqda, yarı nəzarət edilən üsullardan istifadə edilə bilər [Han Jiawei, Kamber Micheline, Pei Jian, 2021].

**Data mining işləmə prosesi.** Data mining üçün sənayelərarası standart proses (CRISP-DM) altı addımlı bir prosesdir və məlumatların istehsalı üçün sənaye standartıdır. Gəlin hər mərhələdə nə gözləyə biləcəyimizi nəzərdən keçirək.

1. Məlumatların toplanması - məlumatların çıxarılması prosesi həll etməyə çalışdığımız problemlə və ya layihə üçün müəyyən bir məqsədlə başlayır. Nəzərə



alınan məqsədləri başa düşmək düzgün, müvafiq məlumat dəstlərinin təhlil etdiyimiz məlumatların olmasını təmin etmək üçün vacibdir. Bu təməl mərhələ təhlil üçün lazımı məlumatların toplanmasını əhatə edir. Bir bank üçün bu, müştəri davranışı və xərcləmə vərdisləri haqqında məlumat əldə etmək üçün müəyyən bir müddət ərzində əməliyyat qeydlərinin toplanması demək ola bilər.

2. Məlumatların anlaşılması - məlumatların tam, dublikatsız və adətən bir neçə mənbədən olmasını təmin edən bütün müvafiq məlumatların toplanmasıdır. Bu, kredit məbləğlərinin bölüşdürülməsinin müşahidəsini və ya kredit kartı istifadəçilərinin orta yaşını kimi demoqrafik yayılmanın başa düşülməsini əhatə edə bilər.

3. Məlumatların hazırlanması - məlumatların hazırlanması zamanı biz məlumatları lazımı formata qoyuruq ki, onlar təhlildən keçməyə və istifadə üçün verilənlər bazasına ötürməyə hazır olsun. Məlumatların hazırlanmasının üç alt addımı ümumi olaraq ETL(extraction, transformation, and loading) adlanan çıxarma, transformasiya və yükləmədir.

4. Modelləşdirmə - modelləşdirmə zamanı verilənlər müxtəlif verilənlərin öyrənilməsi üsulları və alətlərindən keçir, məlumat və məqsədimizdən asılı olaraq düzgün model və ya modellər seçilir. Bu üsullara klasterləşdirmə, reqressiya təhlili və təsnifat daxildir. Bu əsas mərhələdə nümunələri və anlayışları ortaya çıxarmaq üçün dəqiqləşdirilmiş məlumatlar üzərində müvafiq alqoritmlərdən istifadə olunur. Məsələn, bir bank öz müştərilərini kateqoriyalara ayırmaq üçün klasterləşdirmə alqoritmlərindən istifadə edə bilər.

5. Qiymətləndirmə - birinci mərhələdə qoyduğumuz sualın və ya ilkin məqsədin indi cavabı olmalıdır. Ancaq yoxsa, o zaman modelləşdirmənin əvvəlki mərhələsinə qayıdırıq və məlumatlara lazımı düzəlişlər edirik. Nümunələrin çıxarılmasından sonra onlar əhəmiyyətli nəticələr əldə etmək üçün diqqətlə qiymətləndirilir. Bu, aşkar edilmiş nümunənin uyğunluğunun qiymətləndirilməsini və ya nəticələrin real dünya nümunələri ilə təsdiqlənməsini əhatə edir. Məsələn, müəyyən edilmiş tendensiya müəyyən bir ay ərzində kredit borclarının ödənilməməsinin artmasını göstərsə, bank bunu həmin müddət ərzində iqtisadi dalğalanmalar kimi xarici elementlərlə qarşı-qarşıya qoya bilər.

6. Yerləşdirmə - yerləşdirməyə keçməzdən əvvəl, məlumat modelləşdirmə prosesinin bizə hədəfimizə cavab verməsi vacibdir. Bu cavabları əldə etdikdən sonra biz məlumatı strateji olaraq təqdim edə və istifadə edə bilərik [Hand David, Manilla Heikki, Smyth Padhraic, 2001].

**Data mining üsulları.** Məlumat dəstləri daxilində anlayışlar tapmaq üçün müxtəlif strategiyalardan istifadə edəcəyik. Burada bir neçə data mining texnikasına daha yaxından nəzər salaq:

- Klasterləşdirmə - oxşar məlumat nöqtələrinin bir yerdə qruplaşdırılması klasterləşmədir. Nəhayət, klasterləşmədən keçən məlumatlar alt qruplara bölünür. Biz bu alt qruplardan digər verilənlərin öyrənilməsi üsulları üçün giriş məlumatları kimi istifadə edə bilərik.
- Təsnifat - təsnifat həmçinin məlumat dəstlərini daha kiçik qruplara bölür. Təsnifat ümumi, lakin mürəkkəb bir texnikadır, burada model nəticələri proqnozlaşdırmağa kömək etmək üçün oxşarlıqları olan məlumat nöqtələrini axtarır.
- Assosiasiya qaydası - assosiasiya qaydası texnikası bir-biri ilə əlaqəsi olmayan verilənlər dəsti cütü arasında əlaqənin olma ehtimalını göstərir.
- Reqressiya təhlili - reqressiya daha çox riyazi əsaslı texnikadır ki, o, məlumat dəstindəki ən vacib amilləri və onların qarşılıqlı əlaqəsini başa düşməyə kömək edir və nəticədə dəqiq proqnozlar verməyə imkan verir.
- Outlier aşkarlanması - outlier aşkarlanması məlumat dəstindəki potensial səhvləri, eləcə də kənar göstəricini başa düşmək üçün daha yaxından nəzərdən keçirməyə dəyər unikal məlumatları aşkar etmək üçün vacibdir.

Aydın ki, Data Mining metodları tez-tez məlumat dəstlərində nümunələri, əlaqələri və meylləri aşkar etmək üçün istifadə olunur. Bu üsullar böyük məlumat dəstlərindən lazımi məlumat çıxarmaq üçün müxtəlif texnika və alqoritmlərdən istifadə edir. Təhlil və qiymətləndirmə strategiyaları bu metodların məlumatlara təsirini ölçmək və onların nəticələrini şərh etmək üçün istifadə olunur. Tez-tez istifadə olunan bəzi məlumatların öyrənilməsi metodlarının təhlili və qiymətləndirmə strategiyaları bunlardır:

### 1. Qərar ağacları:

- Təhlil strategiyası - qərar ağacları məlumat dəstindəki dəyişənlər arasındakı əlaqələri vizuallaşdırmaq və anlamaq üçün istifadə olunur. Qərar ağacları bir sıra qərar qovşaqlarını və nəticə qovşaqlarını ehtiva edir. O, verilənlər dəstindəki hər bir dəyişənin verilənlərə təsirini və dəyişənlər arasındakı əlaqələri göstərən ağac strukturu yaradır.
- Qiymətləndirmə strategiyası - qərar ağacları dəqiqlik, həssaslıq, spesifiklik göstəricilərindən istifadə etməklə qiymətləndirilir. Bundan əlavə, ağacın ümumiləşdirmə qabiliyyəti çarpaz doğrulama kimi üsullarla qiymətləndirilir.

### 2. Vektor maşınlarına dəstək:

- Təhlil Strategiyası - vektor maşınlarına dəstək təsnifat və reqressiya problemləri üçün istifadə edilən maşın öyrənmə alqoritmidir. Məlumat dəstindəki nümunələri və sinifləri müəyyən etmək üçün istifadə olunur. Bu üsul verilənlər toplusunda nümunələr arasında ən yaxşı ayırmanı tapmağa çalışır.
- Qiymətləndirmə strategiyası - vektor maşınlarına dəstək dəqiqlik, həssaslıq, spesifiklik kimi performans göstəriciləri ilə qiymətləndirilir. Bundan əlavə, təsnifat dəqiqliyi və qərar sərhədinin mürəkkəbliyi kimi amillər də qiymətləndirməyə daxil edilə bilər.

### 3. K-Clustering metodu:

- Təhlil strategiyası - K-clustering verilənlər toplusunda oxşar nümunələri qruplaşdırmaq üçün istifadə olunur. O, məlumat dəstindəki nümunələr arasındakı məsafədən istifadə edərək klaster mərkəzlərini və klasterləri müəyyən edir.
- Qiymətləndirmə strategiyası - K-clustering vasitəsilə, çoxluq daxili homojenlik və çoxluqlararası ayrı-seçkilik kimi daxili qiymətləndirmə göstəricilərindən istifadə etməklə qiymətləndirilir. Bundan əlavə, klasterlərin sayının düzgün müəyyən edilməsi və klaster mərkəzlərinin başlanğıc nöqtələrinin seçilməsi kimi amillər də qiymətləndirmə prosesində mühüm əhəmiyyət kəsb edir.

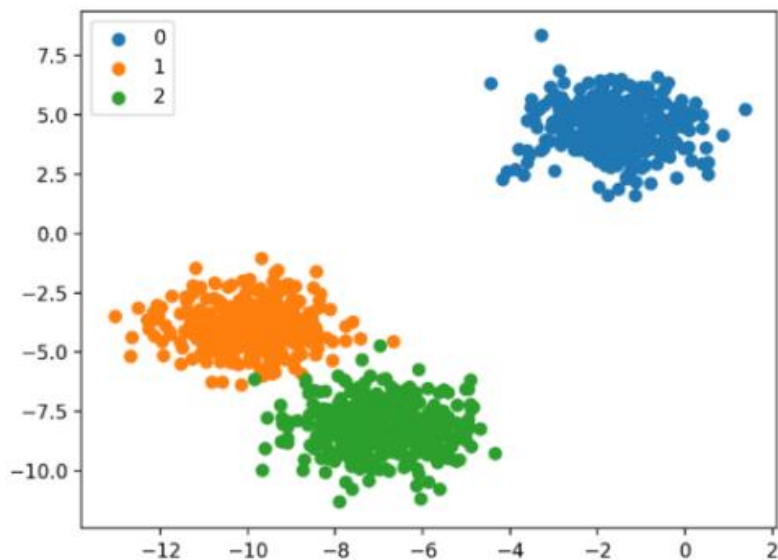
#### 4. Xətti reqressiya:

- Təhlil strategiyası – xətti reqressiya asılı dəyişən ilə bir və ya bir neçə müstəqil dəyişən arasındakı əlaqəni modelləşdirmək üçün istifadə olunur. Bu metoddan verilənlər dəstindəki dəyişənlər arasındakı əlaqəni anlamaq və gələcək dəyərləri proqnozlaşdırmaq üçün istifadə edilə bilər.
- Qiymətləndirmə strategiyası - xətti reqressiya modelləri reqressiya əmsalları, təyinetmə əmsalı (R-kvadrat), orta kvadrat xəta (MSE) və orta mütləq xəta (MAE) kimi metriklərdən istifadə etməklə qiymətləndirilir. Bundan əlavə, modelin mənalılığı və xüsusiyyətlərin əhəmiyyəti kimi amillər də qiymətləndirmə prosesində nəzərə alınır [Witten Hugh, Eibe Frank, 2020].

**Data Mining metodları.** Məlumat elmində və məlumat analitikasında istifadə olunan çox sayda məlumat əldə etmək üsulları mövcuddur. Texnika seçimimiz problemimizin xarakterindən, mövcud məlumatlardan və istədiyiniz nəticələrdən asılıdır. Proqnoz modelləşdirmə data mining-in əsas komponentidir və tarixi məlumat nümunələri əsasında proqnozlar vermək üçün geniş istifadə olunur. Biz həmçinin məlumatlardan hərtərəfli anlayışlar əldə etmək üçün bir sıra üsullardan istifadə edə bilərik. Bu üsullar aşağıdakılardır:

- Klassifikasiya (Classification)

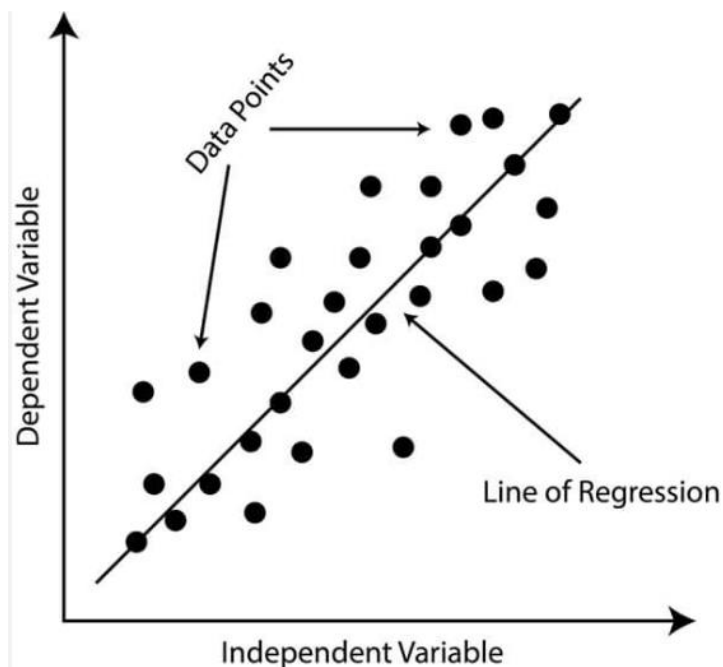
Klassifikasiya məlumat nümunələrinin xüsusiyyətlərinə və ya atributlarına əsaslanaraq, məlumatları əvvəlcədən təyin edilmiş siniflərə və ya kateqoriyalara bölmək üçün istifadə olunan bir texnikadır. Bu, etiketlenmiş məlumatlar üzərində bir model öyrətməyi və yeni, görünməmiş məlumat nümunələrinin sinif etiketlərini proqnozlaşdırmaq üçün istifadə etməyi əhatə edir.



Şək. 2.1. Klassifikasiya

➤ Reqrəssiya (Regression)

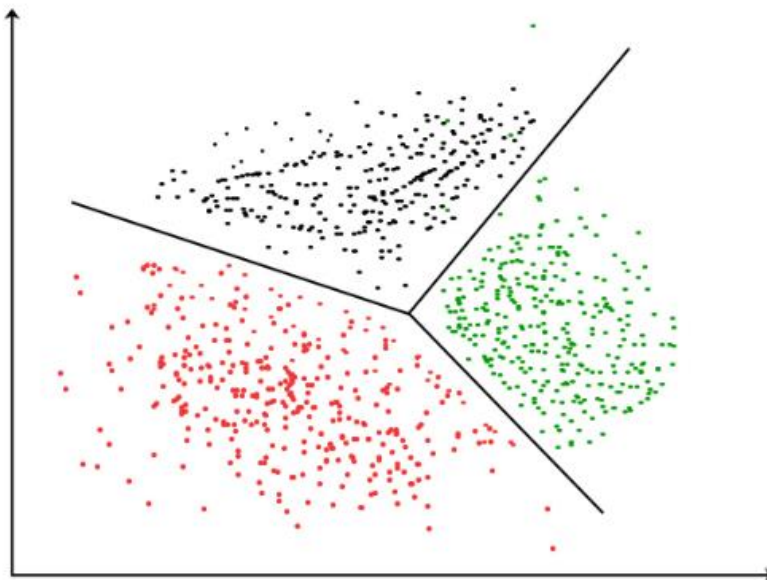
Reqrəssiya, giriş dəyişənləri ilə hədəf dəyişən arasındakı əlaqəyə əsaslanan rəqəmli və ya davamlı dəyərləri proqnozlaşdırmaq üçün istifadə olunur. Dəqiq proqnozlar vermək üçün verilənlərə ən yaxşı uyğun gələn riyazi funksiya və ya modeli tapmaq məqsədi daşıyır.



Şək. 2.2. Reqrəssiya

➤ Klasterləşmə (Clustering)

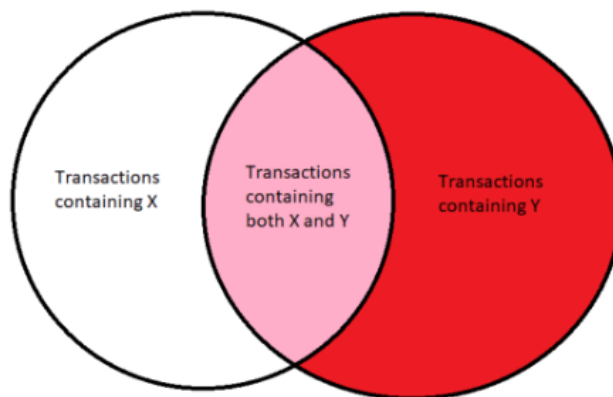
Klasterləşmə oxşar məlumat nümunələrini öz daxili xüsusiyyətlərinə və ya oxşarlıqlarına əsasən qruplaşdırmaq üçün istifadə olunan bir texnikadır. Bu, əvvəlcədən təyin edilmiş siniflər və ya etiketlər olmadan verilənlərdəki orqinal nümunələri və ya strukturları kəşf etməyi hədəfləyir.



Şək 2.3. Klasterləşmə

➤ Assosiasiya qaydası (Association rule)

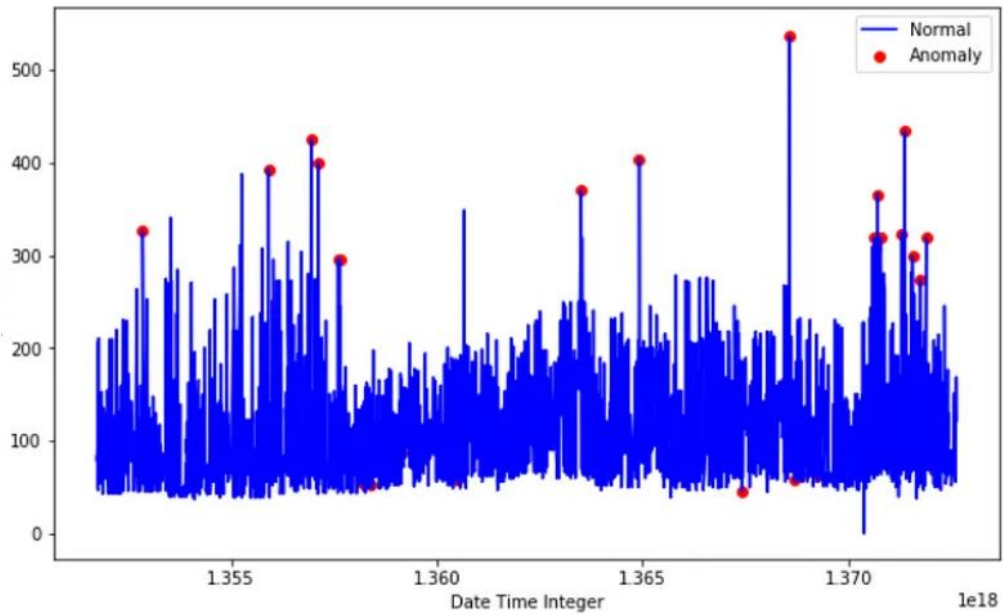
Assosiasiya qaydaları əməliyyat məlumatlarında bir sıra elementlər arasında lazımi əlaqələri və ya nümunələri aşkar etməyə diqqət yetirir. O, tez-tez rast gəlinən elementləri müəyyən etməyə kömək edir və elementlər arasında assosiasiyaları aşkar etmək üçün "əgər X, onda Y" kimi qaydaları yaradır. Bu sadə Venn diaqramı verilənlər dəstinin X və Y element dəstləri arasındakı əlaqəni göstərir.



Şək. 2.4. Assosiasiya qaydası

➤ Anomaliyaların aşkarlanması (Anomaly Detection)

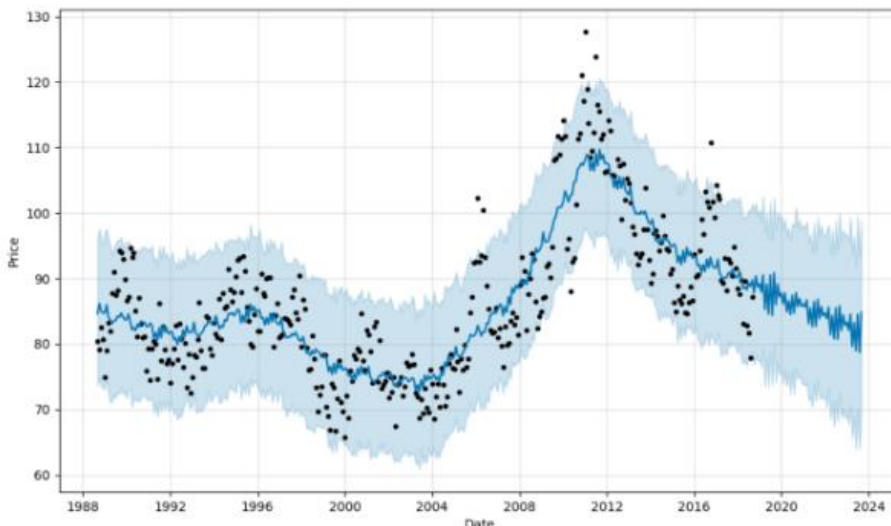
Bəzən kənar analiz adlandırılan anomaliya aşkarlanması gözlənilən nümunələrdən əhəmiyyətli dərəcədə yayınan nadir və ya qeyri-adi məlumat nümunələrini müəyyən etmək məqsədi daşıyır. Bu, saxtakarlıq əməliyyatları, şəbəkə müdaxilələri, istehsal qüsurları və ya hər hansı digər anormal davranışın aşkar edilməsində faydalıdır.



Şək. 2.5. Anomaliyaların aşkarlanması

➤ Zaman seriyalarının təhlili (Time Series Analysis)

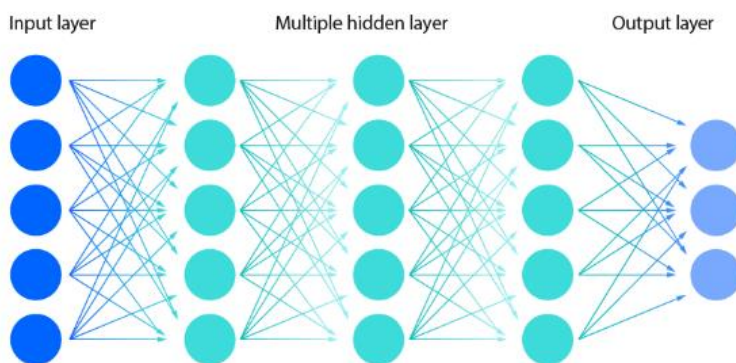
Zaman sıralarının təhlili zamanla toplanmış məlumat nöqtələrinin təhlilinə və proqnozlaşdırılmasına yönəlmişdir. Bu, proqnozlaşdırma, trend təhlili və zamandan asılı verilənlər bazasında anomaliyaların aşkarlanması kimi üsulları əhatə edir.



Şək. 2.6. Zaman seriyalarının təhlili

➤ Neyron şəbəkələr (Neural Networks)

Neyron şəbəkələri insan beyninin strukturu və funksiyasından ilhamlanan bir maşın öyrənməsi və ya AI modelidir. Onlar nümunələri tanımaq, təsnifat, reqressiya və ya digər tapşırıqları yerinə yetirmək üçün verilənlərdən öyrənə bilən bir-biri ilə əlaqəli qovşaqlardan (neyronlar) və təbəqələrdən ibarətdir.



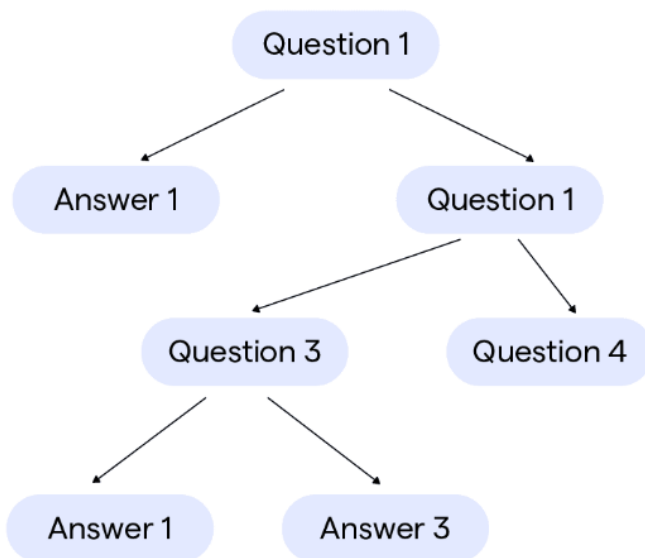
Şək. 2.7. Neyron şəbəkələr

➤ Qərar ağacı (Decision Tree)

Qərar ağacı, əgər/sonra qaydalarından istifadə edərək giriş və çıxış modelləşdirmə əlaqələrinə diqqət yetirən maşın öyrənməsində (ML) məlumatların öyrənilməsi üsuludur. Bu yanaşma ilə biz məlumat girişlərinin çıxışlara necə təsir



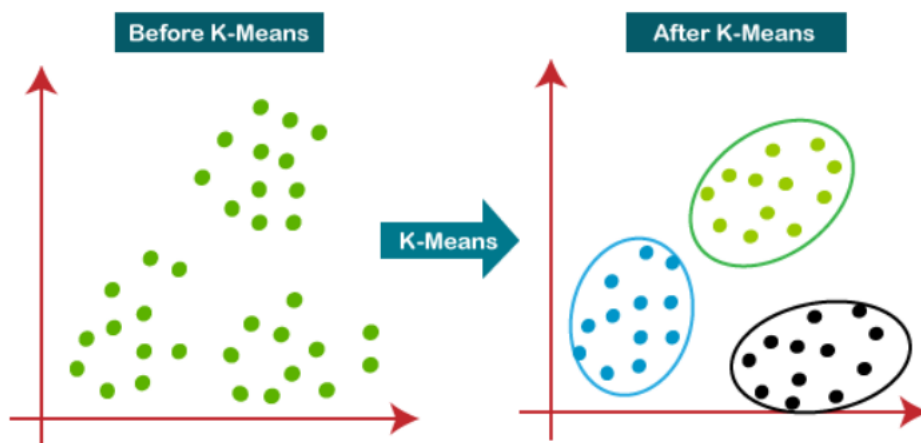
etdiyini öyrənə bilərik. Ağaclar adətən yuxarıdan aşağıya, axın sxeminə bənzər bir quruluşda dizayn edilir.



Şək. 2.7. Qərar ağacı

#### ➤ K-Means

K-Means Clustering, etikətlənməmiş verilənlər toplusunu müxtəlif klasterlərdə qruplaşdırən nəzarətsiz öyrənmə alqoritmidir. Burada K, prosesdə yaradılmalı olan qabaqcadan müəyyən edilmiş klasterlərin sayını müəyyən edir, sanki  $K=2$  olduqda iki çoxluq,  $K=3$  üçün isə üç klaster olacaq və s.

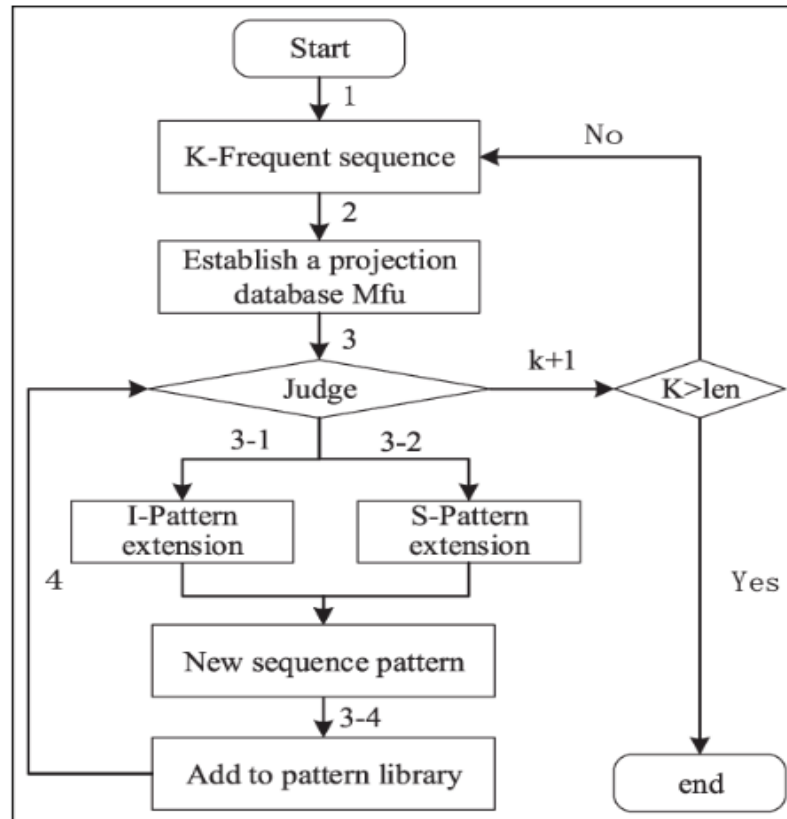


Şək. 2.8. K-Means

#### ➤ Ardıcıl nümunələr (Sequential Patterns)

Bu, ardıcılıqla baş verən nümunələri və ya bir sıra hadisələri kəşf etməyə yönəlmiş bir mədən texnikasıdır. O, əməliyyat məlumatlarının əldə edilməsində geniş

istifadə olunur, lakin çoxsaylı tətbiqlərə malikdir. Məsələn, o, şirkətlərə satışları artırmaq üçün müştərilərə müvafiq maddələr tövsiyə etməyə kömək edə bilər. [Pujari Arun, 2001]



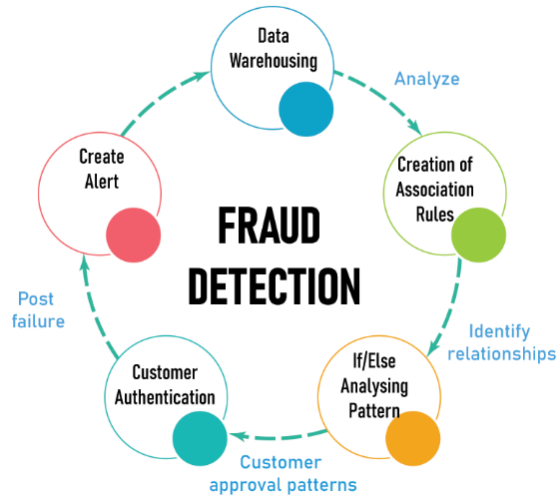
Şək. 2.8. Ardıcıl nümunələr

### 2.3. Bank sektorunda risklərin idarə edilməsi

Data mining böyük və mürəkkəb verilənlər bazasından lazımi məlumatların çıxarılması prosesidir. O, qərar qəbulunu və performansını yaxşılaşdırmağa kömək edə biləcək nümunələri, meylləri və əlaqələri aşkar etmək üçün istifadə edilə bilər. Data mining tətbiqlərindən biri fraud dedection, yəni saxtakarlığın aşkarlanmasıdır.

**Saxtakarlığın aşkarlanması.** Saxtakarlığın aşkarlanması saxtakarların saxta yollarla pul və ya əmlak əldə etmələrini aşkara çıxaran və bunun qarşısını alan bir prosesdir. Bu, saxtakarların dolayı yollar ilə pul və ya əmlak əldə etmək cəhdinin aşkar edilməsi və qarşısının alınması üçün həyata keçirilən tədbirlər toplusudur.

Saxtakarlığın aşkarlanması bank, tibb, sığorta, hökumət və ictimai sektorlarda və s geniş yayılmışdır.



Şək. 2.9. Saxtakarlığın aşkarlanmasının işləmə prinsipi

Bank işində saxtakarlığın aşkarlanması prosesi, maliyyə institutları üçün risklər yaradan bir sıra aldadıcı təcrübələrin qarşısını almaq üçün strateji olaraq hazırlanmış mürəkkəb və dinamik bir sistemdir. Eyni zamanda, saxtakarlıq fəaliyyətləri ilə bağlı çoxşaxəli riskləri müəyyən etmək, qiymətləndirmək və azaltmaq üçün dəqiq şəkildə inteqrasiya olunmuş qabaqlayıcı tədbirlər və qabaqcıl texnologiyalardan ibarətdir.

Maliyyə institutları getdikcə daha çox bu cür saxtakarlığın qarşısını ala bilmək üçün həm əl ilə, həm də avtomatlaşdırılmış üsullardan istifadə edirlər. Tranzaksiyaların monitorinqi və anomaliyaların aşkarlanmasından tutmuş davranış təhlilinə qədər əsas məqsəd potensial saxtakarlığın göstəricisi olan pozuntuları qısa bir müddətdə müəyyən etməkdir. Bu proses isə tək bir yanaşma ilə məhdudlaşmır; həm taktikaların, həm də texnologiyaların davamlı inkişafını əhatə edən hərtərəfli strategiyanı özündə ehtiva edir.

Aldadıcı təcrübələr spektrinə hesabın ələ keçirilməsi, kredit kartı saxtakarları və maliyyə sistemindəki zəifliklərdən istifadə edən digər çirkin fəaliyyətlər daxildir, lakin bununla da kifayətlənmir və bu manevrlər tez-tez təhlükəsizlik protokollarındakı boşluqlardan istifadə edərək, saxtakarlığın aşkarlanmasına diqqətli və uyğunlaşan yanaşma tələb edir.

Digər mühüm bir cəhət isə bankın reputasiyasının qorunub saxlanması dələduzluğun aşkarlanmasından təsirlənir. Saxtakarlıq fəaliyyətinin bankın imicinə vurduğu zərər uzunmüddətli və geniş ola bilər. Müştərilər öz təhlükəsizliyinə aktiv bağlılıq nümayiş etdirən qurumlara daha çox güvənirlər. Güclü saxtakarlıq aşkarlama sistemi təkcə müştəriləri qorumur, həmçinin bankın reputasiyasını tez-tez maliyyə qalmaqalları ilə müşayiət olunan aşınmadan qoruyur.

Saxtakarlığın aşkarlanması ümumiyyətlə məlumatların təhlilinə əsaslanan üsulları ehtiva (əhatə) edir. Bu üsullar geniş şəkildə statistik məlumatların təhlili üsulları və süni intellekt əsaslı üsullar kimi təsnif edilir. Hər ikisi üçün geniş şəkildə izah aşağıdakı kimidir:

- Statistik məlumatların təhlili üsulları - saxtakarlığın aşkarlanması üçün statistik məlumatların təhlili təfərrüatlı araşdırmalar aparmaqla saxtakarlıq məlumatlarının toplanması, saxtakarlığın aşkarlanması və saxtakarlığın təsdiqi kimi müxtəlif statistik əməliyyatları həyata keçirir. Bu texnikalar aşağıdakı növlərə bölünür:

1. Statistik parametrlərin hesablanması - məlumatların tutulması prosesi zamanı toplanmış saxtakarlıqla bağlı məlumatlar üçün orta göstəricilər, performans göstəriciləri, kvantillər və ehtimal paylamaları kimi müxtəlif statistik parametrlərin hesablanmasına aiddir.
2. Reqressiya təhlili - maraq doğuran iki və ya daha çox dəyişən arasındakı əlaqəni yoxlamağa imkan verir və həmçinin müstəqil və asılı dəyişənlər arasındakı əlaqəni qiymətləndirir.
3. Ehtimal paylamaları və modelləri - müxtəlif parametrlər və ya ehtimal paylamaları baxımından müxtəlif biznes saxtakarlıq fəaliyyətlərinin modelləri və ehtimal paylamaları müəyyənləşdirilir.
4. Məlumatların uyğunlaşdırılması - iki toplanmış məlumat dəstini (yəni, saxtakarlıq məlumatları) müqayisə etmək üçün istifadə olunur. Proses ya alqoritmlər, ya da proqramlaşdırılmış döngələr əsasında həyata keçirilə bilər. Bundan başqa, məlumatların uyğunlaşdırılması dublikat qeydləri

silmək və marketinq, təhlükəsizlik və ya digər məqsədlər üçün iki məlumat dəsti arasında əlaqələri müəyyən etmək üçün də istifadə olunur.

- Süni intellektə əsaslanan üsullar - saxtakarlığın qarşısının alınması üçün AI-nin tətbiqi şirkətlərə daxili təhlükəsizliyini artırmaq və biznes proseslərini asanlaşdırmaqda böyük rol oynayır. Təkmilləşdirilmiş səmərəlilik sayəsində süni intellekt maliyyə institutlarında saxtakarlığın qarşısını almaq üçün vacib bir texnologiya kimi ortaya çıxdı. Süni intellektə əsaslanan saxtakarlığın aşkarlanması üsulları aşağıdakı növlərə bölünür:

1. Saxtakarlığın aşkarlanması və qarşısının alınması üçün verilənlərin əldə edilməsi məlumatı təsnif edir, qruplaşdırır və seqmentlərə ayırır və məlumatlarda maraqlı nümunələri, o cümlədən saxtakarlıqla bağlı olanları ifadə edə bilən assosiasiyaları və qaydaları avtomatik tapır.
2. Neyron şəbəkələri - daxili auditlərdə və ya rəsmi maliyyə sənədlərində əldə edilən nəticələrlə müqayisə oluna bilən saxtakarlıqla bağlı məlumatların təsnifatını, qruplaşdırılmasını, ümumiləşdirilməsini və proqnozlaşdırılmasını həyata keçirir.
3. Maşın öyrənməsi (ML) - ML alqoritmlərinin tarixi saxtakarlıq nümunələrindən öyrənmək və gələcək əməliyyatlarda onları tanımaq qabiliyyəti sayəsində mümkün olur. Maşın öyrənməsi nəzarət edilən və ya nəzarətsiz öyrənmə metodlarından istifadə edir. Nəzarət olunan öyrənmədə bütün qeydlərin təsadüfi alt nümunəsi əl ilə ya “saxtakar” və ya “saxtakar olmayan” kimi təsnif edilir. Nəzarətsiz öyrənmədə isə metodlar ümumi nümunələri (yəni, saxtakarlıq) və xam məlumatda korrelyasiyaları axtarır və proqnozlar əlavə etiketləmə olmadan qurulur.
4. Nümunənin tanınması - alqoritmləri avtomatik (yəni, nəzarətsiz) və ya əl ilə (yəni, nəzarət altında) təxmini sinifləri, qrupları və ya şübhəli davranış nümunələrini aşkarlayır.

Saxtakarlığın aşkarlanmasının effektivliyi ən yaxşı təcrübələrə əsaslanır. Saxtakarlığın qarşısının alınması çərçivəsinin həyata keçirilməsi davamlı monitorinqi,

müntəzəm risk qiymətləndirmələrini və inkişaf edən təhdidləri qabaqlamaq üçün adaptiv strategiyaları əhatə edir.

1. Real vaxt rejimində monitoring - yalnız müştərini tanı standartlarına (Know Your Customer standards) və çirkli pulların yuyulmasına qarşı qanunlara (Anti-Money Laundering laws) uyğunluq tədbiri kimi deyil, həm də saxtakarlıq fəaliyyətinin aşkar edilməsi üçün güclü strategiya kimi xidmət edən kritik təcrübədir.

2. Risklərin müntəzəm qiymətləndirilməsi – bu daha çox, saxtakarlıqda yaranan tendensiyaların qiymətləndirilməsini, mövcud sistemlərdə zəifliklərin qiymətləndirilməsini və müvafiq olaraq strategiyaların uyğunlaşdırılmasını əhatə edir.

3. Adaptiv strategiyalar - saxtakarlığın aşkarlanması sahəsində çeviklik çox önəmlidir. Yaranan təhlükələrlə yanaşı inkişaf edən adaptiv strategiyaların həyata keçirilməsi bankların bir addım öndə qalmasını təmin edir. Bu, yeni nümunələrdən öyrənən və aşkarlama imkanlarını davamlı olaraq təkmilləşdirən maşın öyrənmə alqoritmlərinin inteqrasiyasını əhatə edə bilər.

4. Müştərilərin maarifləndirilməsi və məlumatlandırılması - müştəriləri ən son saxtakarlıq sxemləri haqqında maarifləndirmək və ən yaxşı təhlükəsizlik təcrübələri haqqında məlumatlılığı təşviq etmək onlara saxtakarlığın qarşısının alınmasında fəal rol oynamağa imkan verir. Banklar ümumi saxtakarlar haqqında müntəzəm yeniliklər təqdim edə və təhlükəsiz bankçılıq təcrübələrinə dair təlimatlar təklif edə bilərlər.

5. Çox qatlı doğrulama - buraya biometrik doğrulama, iki faktorlu autentifikasiya və ya hesablara girişin yalnız səlahiyyətli şəxslərə verilməsini təmin etmək üçün digər qabaqcıl üsullar daxil ola bilər.

Saxtakarlığın aşkarlanması və qarşısının alınması istənilən biznes üçün əsas prioritet olmalıdır. Yaxşı dizayn edilmiş və tətbiq edilmiş saxtakarlığın aşkarlanması sistemi təşkilat daxilində baş verə biləcək potensial saxtakarlıq şanslarını əhəmiyyətli dərəcədə azalda bilər. Bundan əlavə, dələduzluğun vaxtında aşkarlanması gələcək potensial itkiləri azaltmaqla biznesə müsbət təsir göstərir. Süni intellekt və statistik məlumatların təhlili kimi effektiv aşkarlama üsulları potensial saxtakarlara qarşı maneə rolunu oynayır. Tənzimləyici tələblər və uyğunluq tələbləri artdıqca, saxtakarlığın

aşkarlanması və qarşısının alınması proqramının həyata keçirilməsi son dərəcə vacib olmuşdur [Husain Mohd, Khan Mohammad, 2020].

**Müştəri əlaqələrinin idarə edilməsi.** CRM (Customer Relationship Management) - müştəri əlaqələrinin idarə edilməsi deməkdir. Bu, müştərilərlə bank arasında əlaqələrin inkişaf etdirilməsi prosesidir. Ən yaxşı texnologiyayı gətirərək müştəri məmnuniyyətini artırmağa önəm verir. CRM əsas məqsədi mükəmməl xidmət göstərməklə uzunmüddətli müştəri əlaqələri yaratmaqdır.

CRM texnologiyasındakı bütün irəliləyişlər üçün, düzgün idarəetmə olmadan, bir CRM sistemi müştəri məlumatlarının saxlandığı təriflənmiş verilənlər bazasından sadəcə bir az çox ola bilər. Məlumat dəstləri qoşulmalı, paylanmalı və təşkil edilməlidir ki, istifadəçilər lazım olan məlumatı rahatlıqla əldə edə bilsinlər. Əgər məlumat dəstləri tək nəzarət paneli və ya interfeysdə birləşdirilməyibsə və təşkil edilməyibsə, şirkətlər müştərinin vahid görünüşünə nail olmaq üçün mübarizə apara bilər. Sistemlərdə təkrarlanan müştəri məlumatı və ya köhnəlmiş məlumat olduqda müştərini izləməkdə çətinliklər yaranır. Bu problemlər, telefon zəngləri zamanı uzun gözləmə müddətləri və texniki dəstək işlərinin düzgün aparılmaması kimi problemlər səbəbindən müştəri təcrübəsinin azalmasına səbəb ola bilər. Şirkətlər CRM məlumatlarını xarici məlumat mənbələri ilə əlavə etməzdən əvvəl təkrarlanan və ya natamam qeydləri aradan qaldırmaq üçün mövcud müştəri məlumatlarını təmizlədikdə CRM sistemləri daha yaxşı işləyir.

Çox vaxt bankirlər CRM-in nə demək olduğunu və onun necə işlədiyini başa düşürlər. Buna baxmayaraq, heç də hamı CRM-in bank işində əhəmiyyətini dərk etmir və belə sistemləri öz bizneslərinə inteqrasiya etmir. Görünür ki, bank CRM olmadan da gəlirli ola bilər, lakin bank özünü çaşdırmağa bilər, çünki müştəri yönümlü məhsullar qarşılıqlı əlaqə prosesi ilə bağlı hər şeyi əhəmiyyətli dərəcədə artırır. CRM-in üç əsas xidmətini qeyd edə bilərik:

1. Potensial və mövcud müştəriləri saxlamaq - güclü bir promo kampaniyadan sonra müştərini laqeyd etmək problem olardı. CRM istifadəçilərdən gələn

telefon zəngləri, e-poçtlar və digər sorğular kimi bütün potensial müştəriləri izləməyi və yadda saxlamağa kömək edir.

2. İşçilərə nəzarət etmək və standartları müəyyən etmək - CRM olmayan banklardakı işçilər Excel kimi müxtəlif mühasibat üsulları və ya hətta öz yaddaşları ilə işləri görsələr də, hamısı bir yerdə həllər xaosdan xilas olur və hər şeyi birləşdirir. Bu isə performansını artırır.
3. Məlumat toplamaq və onları birləşdirmək - idarəetmə alətləri əlaqə və sifarişlər kimi vacib məlumatları birləşdirən vahid verilənlər bazası yaradır. CRM ilə satış prosesini daha dəqiq təhlil etmək və planlaşdırmaq mümkündür.

CRM sistemlərinin dörd əsas təchizatçıları Microsoft, Oracle, Salesforce və SAP-dır. Bu dörd sistem böyük şirkətlər üçün ən yaxşı sistemdir, digər provayderlər kiçik və orta biznes arasında məşhurdur. Təklif olunan CRM texnologiyasının aşağıdakı növləri vardır:

- Bulud əsaslı CRM - bulud hesablamalarından istifadə edən CRM həm də xidmət və ya tələb olunan proqram təminatı kimi də tanınır. Məlumat işçilərin istənilən vaxt və internet bağlantısı olan hər yerdə daxil ola biləcəyi xarici, uzaq şəbəkədə saxlanılır. Bəzi hallarda, üçüncü tərəf xidmət təminatçısı sistemin quraşdırılmasına və saxlanmasına nəzarət edir. Buludun sürətli və asan yerləşdirmə imkanları məhdud texnoloji təcrübə və ya resursları olan şirkətləri cəlb edir. Məlumat təhlükəsizliyi bulud əsaslı sistemlərdən istifadə edən şirkətlər üçün müəyyən qədər narahatlıq doğurur, bu səbəb şirkət öz məlumatlarının saxlanmasına fiziki nəzarət etmir. Əgər bulud provayderi fəaliyyətini dayandırarsa və ya başqa şirkət tərəfindən alınarsa, müəssisənin məlumatları pozula və ya itirilə bilər. Uyğunluq problemləri həm də məlumat ilkin olaraq şirkətin daxili sistemindən bulud sisteminə köçürüldükdə yarana bilər. Şirkətlər bulud CRM-ni qənaətcil bir seçim kimi nəzərdən keçirə bilər. Satıcılar adətən abunə əsasında istifadəçidən müəyyən ödəniş alır və aylıq və ya illik ödəniş seçimi təklif edirlər. Bununla belə, xərc hələ də narahatlıq doğura bilər, çünki proqram təminatı üçün abunə haqqının ödənilməsi zamanla yerli modelə



investisiya qoymaqdan daha baha başa gələ bilər. Daha populyar olan bulud əsaslı CRM provayderlərinə HubSpot, Salesforce, Zendesk və Zoho daxildir.

- Yerli CRM - CRM proqram təminatından istifadə edən şirkətə verilənlər bazası və məlumatların idarə edilməsi, nəzarəti, təhlükəsizliyi və saxlanması vəzifəsini qoyur. Bu yanaşma ilə şirkət bulud CRM provayderindən illik abunəliklər almaq əvəzinə lisenziyaları əvvəlcədən alır. Proqram təminatı şirkətin serverlərində yerləşir və istifadəçi istənilən təkmilləşdirmənin qiymətini öz üzərinə götürür. Həmçinin, adətən şirkətin məlumatlarını tam inteqrasiya etmək üçün uzunmüddətli quraşdırma prosesi tələb olunur. Kompleks CRM ehtiyacları olan şirkətlər yerli yerləşdirmədən yararlanmağa bilər. Apeian və Salesforce kimi bir çox bulud əsaslı provayderlər də CRM proqram təminatının yerli versiyalarını təklif edirlər.
- Açıq mənbəli CRM – bu sistem mənbə kodunu ictimaiyyətə təqdim edərək sistemdən istifadə edən şirkətə heç bir ödəniş etmədən dəyişiklik etməyə imkan verir. Açıq mənbəli CRM sistemləri sosial media kanallarında məlumat bağlantılarının əlavə edilməsinə və fərdiləşdirilməsinə imkan verir, sosial CRM təcrübələrini inkişaf etdirmək istəyən şirkətlərə kömək edir. Bitrix24, OroCRM, SugarCRM və SuiteCRM kimi platformalar Salesforce, Microsoft və digər təchizatçıların özəl platformalarına alternativlər təklif edir. Bu CRM yerləşdirmə üsullarından hər hansı birinin qəbul edilməsi şirkətin biznes ehtiyaclarından, resurslarından və məqsədlərindən asılıdır, çünki hər birinin onunla bağlı müxtəlif xərcləri var.

**CRM-in banklara köməkliyi.** CRM müştərinin saxlanmasını artırır. Bu, banklara müştərilərini diqqətlə izləməyə və nə istədiklərini bilməyə imkan verir. Artıq bankdan razı olmayan və digər banklara üz tutan müştəriləri müəyyən etməyə kömək edir. Banklar daha sonra həmin müştəriləri saxlamaq üçün lazımi addımları atmağa bilər.

Müştəri saxlamaq və çarpaz satış imkanlarını artırmaqla gəlir artımına kömək edir. Bu, müştəriləri bankla bizneslərini inkişaf etdirməyə və bank məhsulları və xidmətlərinə olan xərclərini artırmağa sövq edir. Həmçinin, müştərilərin bankdan ehtiyac duymağa yeni məhsullar əlavə etməklə gəlirləri artırmağa kömək edir.

Müştəri davranışı və üstünlükləri haqqında daha yaxşı təsəvvürlər verir. Bu, hər bir müştərinin davranışını anlamağa kömək edir və bununla da bankların müştərilərinə fərdiləşdirilmiş xidmət təklif etməsini asanlaşdırır. Tələbləri və ehtiyacları üçün paketlər təqdim etməklə onlara daha yaxşı xidmət göstərə bilirlər.

Müştəri təcrübəsini yaxşılaşdırır. Müştərilər bankdan istədiklərini o zaman əldə edirlər ki, bank onları CRM vasitəsilə kifayət qədər yaxşı başa düşür. Bu, bank işçiləri ilə onların müştəriləri arasında ünsiyyəti asanlaşdırır, müştəri təcrübəsini yaxşılaşdırır.

CRM, banklara gəlir əldə etmək imkanlarını genişləndirməkdə mühüm rol oynayır. O, yeni və ya mövcud müştərilər tərəfindən tələb olunan yeni xidmətləri müəyyən etməklə banklara gəlir yaratmaq imkanlarını təkmilləşdirməyə kömək edir və bununla da banklar üçün daha çox gəlir əldə edir. Həmçinin, bu, banklara mövcud və potensial müştərilərlə yeni iş imkanlarını araşdırmaq imkanı verir ki, bununla da banklar üçün daha çox gəlir əldə edir. [Linoff Gordon, Beryy Michael, 2021].

**Risk idarəetməsi və kredit skoringi.** Data mining tətbiqlərindən biri kredit skoringidir ki, bu da borcalanın keçmiş və indiki maliyyə davranışlarına əsaslanaraq kredit qabiliyyətinin qiymətləndirilməsidir. Kredit skoringi həm kreditorlar, həm də borcalanlar üçün vacibdir, çünki bu, kreditin mövcudluğuna, dəyərinə və şərtlərinə təsir göstərir.

Kredit skoringi yaxşılaşdırmaq üçün istifadə edilə bilən müxtəlif məlumatların öyrənilməsi üsulları var, məsələn:

1. Təsnifat - bu texnika hər bir borcalana kredit profilinə əsasən etiket və ya kateqoriya təyin etməyi nəzərdə tutur. Məsələn, borcalan ödəniş tarixinə, gəlirinə, borcuna və digər amillərə görə yaxşı, pis və ya neytral kimi təsnif edilə bilər. Təsnifat kreditorlara hər bir borcalanın risk səviyyəsini müəyyən etməyə və onlara müvafiq kredit məhsulları və faiz dərəcələri təklif etməyə kömək edə bilər. Təsnifat həm də borcalanlara kredit statuslarını anlamağa və onu yaxşılaşdırmaq üçün tədbirlər görməyə kömək edə bilər. Ümumi təsnifat üsullarından bəziləri logistik reqressiya, qərar ağacları, neyron şəbəkələri və dəstək vektor maşınlarıdır.

2. Klasterləşdirmə - bu texnika oxşar xüsusiyyətlərə və ya davranışlara malik borcalanların qruplara bölünməsinə nəzərdə tutur. Məsələn, borcalanların yaşı, cinsi, məşğuliyyəti və ya yerləşdiyi yer əsasında klaster yaradıla bilər. Klasterləşmə kreditlərə öz müştəri bazalarını segmentləşdirməyə və marketing və xidmət strategiyalarını buna uyğun olaraq uyğunlaşdırmağa kömək edə bilər. Klasterləşdirmə həm də borcalanlara özlərini həmyaşıdları ilə müqayisə etməyə və onların ən yaxşı təcrübələrini öyrənməyə kömək edə bilər. Ümumi klasterləşdirmə üsullarından bəziləri k-orta, iyerarxik və sıxlığa əsaslanan klasterləşmədir.
3. Assosiasiya qaydalarının öyrənilməsi - bu texnika müxtəlif dəyişənlər və ya maddələr arasında əlaqələri və ya assosiasiyaları təsvir edən qaydaların tapılmasını əhatə edir. Məsələn, bir qayda tapmaq olar ki, gəliri yüksək və borcun gəlirə nisbəti aşağı olan borcalanların kreditlərini vaxtında qaytarmaq ehtimalı daha yüksəkdir. Assosiasiya qaydaları kreditlərə borcalanların kredit davranışına təsir edən amilləri aşkar etməyə və effektiv kredit siyasəti və təşviqləri hazırlamağa kömək edə bilər. Assosiasiya qaydaları həmçinin borcalanlara kredit hesablarına təsir edən amilləri müəyyən etməyə və vərdişlərini müvafiq olaraq dəyişdirməyə kömək edə bilər. Ümumi birləşmə qaydaları üsullarından bəziləri Apriori, FP-böyümə və Eclatdır.
4. Anomaliyaların aşkarlanması - bu texnika normal və ya gözlənilən modeldən kənara çıxan kənar göstəricilərin və ya anormal halların aşkar edilməsini əhatə edir. Məsələn, anomaliya borcalan qəflətən xərcləmə və ya ödəmə davranışını dəyişdikdə və ya borcalanın kredit profili demoqrafik və ya sosial atributlarına uyğun gəlmədikdə aşkar edilə bilər. Anomaliyaların aşkarlanması kreditlərə saxtakarlıq və ya defolt riskləri aşkar etməyə və qabaqlayıcı və ya düzəldici tədbirlər görməyə kömək edə bilər. Anomaliyaların aşkarlanması həmçinin borcalanlara kredit fəaliyyətlərinə nəzarət etməyə və hər hansı şübhəli və ya səhv əməliyyatlar barədə məlumat verməyə kömək edə bilər. Ümumi anomaliya aşkarlama üsullarından bəziləri statistik, məsafəyə əsaslanan və sıxlığa əsaslanan üsullardır.

5. Reqressiya - bu texnika asılı dəyişən ilə verilənlər bazasında bir və ya bir neçə müstəqil dəyişən arasındakı əlaqəni ən yaxşı şəkildə təsvir edən riyazi funksiya və ya tənliyin tapılmasını nəzərdə tutur. Məsələn, cari kredit davranışına və digər amillərə əsaslanaraq gələcək kredit xallarını proqnozlaşdırmaq üçün reqressiyadan istifadə edə və sonra öz hərəkətlərini müvafiq olaraq tənzimləyə bilər. Bunu etməklə, əvvəldən plan qura və maliyyə qərarlarının kredit balları ilə bağlı nəticələrini təxmin edə bilər. Məsələn, kredit kartı balansını hər ay tam ödəsələr, altı ay ərzində kredit ballarının 10 bal artacağını və sonra ardıcıl olaraq bu plana əməl edəcəyini öyrənmək üçün reqressiyadan istifadə etmək olar.

Bunlar kredit skoringi yaxşılaşdırmaq üçün istifadə edilə bilən bəzi məlumatların öyrənilməsi üsullarıdır. Bununla belə, həll edilməli olan bəzi problemlər və məhdudiyyətlər var, məsələn:

- Məlumatların keyfiyyəti - kredit skoringi üçün istifadə olunan məlumatlar dəqiq, tam, ardıcıl və vaxtında olmalıdır. Zəif məlumat keyfiyyəti qeyri-dəqiq və ya yanlış nəticələrə gətirib çıxara bilər və kredit qiymətləndirmə modellərinin etibarlılığına və etibarlılığına təsir göstərə bilər.
- Məlumatların məxfiliyi - kredit skoringi üçün istifadə edilən məlumatlar icazəsiz girişdən, istifadədən və ya açıqlamadan qorunmalıdır. Məlumatların məxfiliyi hakerlər, insayderlər və ya məlumatları zərərli və ya qeyri-etik məqsədlər üçün sui-istifadə edə bilən üçüncü tərəflər tərəfindən pozula bilər.
- Məlumat etikası - kredit skoringi üçün istifadə olunan məlumatlar ədalətli, şəffaf və hesabatlı olmalıdır. Məlumat etikası borcalanların və ya borc verənlərin maraqlarına və ya hüquqlarına zərər verə biləcək qərəzli yanaşma, ayrı-seçkilik və ya manipulyasiya ilə pozula bilər.

Buna görə də, kredit skoringi üçün məlumatların öyrənilməsi üsulları ehtiyatla və məsuliyyətlə, hüquqi və etik standart və qaydalara uyğun olaraq tətbiq edilməlidir. Data mining kredit reytingini yaxşılaşdırmaq üçün güclü bir vasitə ola bilər, lakin o, insan mühakiməsi və ekspertizasının əvəzi kimi istifadə edilməməlidir.

**Kredit qiymətləndirmə modelinin komponentləri.** Kredit qiymətləndirmə modeli fiziki şəxslərin, müəssisələrin və ya maliyyə alətlərinin kredit qabiliyyətini qiymətləndirmək üçün kreditorlar, maliyyə institutları və ya kredit agentlikləri tərəfindən istifadə olunan alət və ya sistemdir. Əsas məqsədi borcalanın borc öhdəliklərini yerinə yetirməməsi ehtimalını qiymətləndirməkdir. Bu modelin əsas komponentləri aşağıdakılardır:

- Proqnozlaşdırılan dəyişənlər - bunlar kredit riskini proqnozlaşdırmaq üçün istifadə edilən giriş dəyişənləridir. Bunlara borcalanın xüsusiyyətləri (məsələn, yaş, gəlir, məşğulluq statusu), kredit tarixçəsi (məsələn, kredit hesabı, ödəniş tarixçəsi) və kreditə xas amillər (məsələn, kredit məbləği, kreditin dəyər nisbəti) daxil ola bilər.
- Qiymətləndirmə alqoritmi - qiymətləndirmə alqoritmi proqnozlaşdırılan dəyişənlərə əsaslanaraq kredit xalını hesablayır. Ümumi alqoritmlərə logistik reqressiya, qərar ağacları, neyron şəbəkələri və ansambl metodları daxildir. Hər bir alqoritmin öz güclü tərəfləri və məhdudiyyətləri var və alqoritmin seçimi kreditorun konkret tələblərindən və məqsədlərindən asılıdır.
- Kredit skoru - kredit balı kredit skoring modelinin nəticəsidir və borcalanın kredit qabiliyyətini təmsil edir. Bu adətən bir neçə yüzdən bir neçə minə qədər dəyişən ədədi dəyərdir, yüksək ballar aşağı kredit riskini, aşağı ballar isə daha yüksək riski göstərir.

**Kredit qiymətləndirmə modelinin yaradılma prosesi.** Kredit qiymətləndirmə modelinin yaradılması prosesi boyunca biznes məqsədlərinə və tənzimləyici tələblərə uyğunluğu təmin etmək üçün vacibdir. Müntəzəm yoxlama, monitoring və texniki xidmət də modelin uzunmüddətli uğuru üçün çox vacibdir.

1. Məlumatların toplanması - kredit skoringi modelinin inkişaf prosesi borcalanın xüsusiyyətləri, kredit tarixçəsi və kredit fəaliyyəti haqqında tarixi məlumatların toplanması ilə başlayır. Bu məlumatlar müxtəlif daxili və xarici mənbələrdən əldə edilir, o cümlədən:

- Daxili mənbələr - kredit müraciətləri, müştəri məlumat bazaları və kredit təşkilatı tərəfindən aparılan əməliyyat qeydləri.
  - Xarici mənbələr - kredit hesabatları və xalları, habelə iflaslar, girovlar və qərarlar haqqında məlumatları ehtiva edən ictimai qeydləri təqdim edən kredit büroları. Modelin möhkəmliyini və ümumiləşdirilməsini təmin etmək üçün toplanmış məlumatlar müxtəlif borcalanları və kreditləri əhatə etməlidir.
2. Məlumatların əvvəlcədən emalı - Məlumat toplandıqdan sonra keyfiyyətini və modelləşdirmə üçün uyğunluğunu təmin etmək üçün ilkin emaldan keçir. Bu, bir neçə addımı əhatə edir, o cümlədən:
- Təmizləmə - məlumatlardakı hər hansı qeyri-dəqiqlik, uyğunsuzluq və ya çatışmayan dəyərlərin aradan qaldırılması və ya düzəldilməsi.
  - Transformasiya - kateqorik dəyişənlərin ədədi formatlara kodlaşdırılması və ədədi dəyişənlərin standartlaşdırılmış diapazona miqyaslanması.
3. Dəyişən seçimi - sonra, proqnozlaşdırılan dəyişənlər kredit riskinin qiymətləndirilməsində onların uyğunluğu və proqnozlaşdırma gücü əsasında seçilir. Bu proses daxildir:
- Məlumat təhlili - ən informativ xüsusiyyətləri müəyyən etmək üçün potensial dəyişənlər və kreditin defolt dərəcələri arasındakı əlaqələrin təhlili.
  - Statistik üsullar - dəyişənlərin əhəmiyyətini qiymətləndirmək üçün korrelyasiya təhlili, chi-square tests və ya məlumat əldə etmək kimi statistik metodlardan istifadə. Məqsəd kredit qabiliyyətinin əsas determinantlarını effektiv şəkildə ələ keçirən dəyişənlərin alt çoxluğunu müəyyən etməkdir.
4. Modelin inkişafı - seçilmiş dəyişənlərlə kredit qiymətləndirmə modelini hazırlamaq üçün müxtəlif modelləşdirmə üsulları tətbiq edilir. Ümumi istifadə edilən alqoritmlərə aşağıdakılar daxildir:
- Logistik reqressiya - kreditin defolt və ya qeyri-defolt kimi ikili nəticənin ehtimalını qiymətləndirmək üçün istifadə olunan xətti model.

- Qərar ağacları - verilənləri proqnozlaşdıran dəyişənlərə əsaslanan alt çoxluqlara bölən ağac əsaslı modellər, qərar qaydalarının ağaca bənzər strukturuna gətirib çıxarır.
- Neyron şəbəkələr - insan beyninin strukturu və funksiyasından ilhamlanan, verilənlərdən mürəkkəb nümunələri öyrənməyə qadir olan kompleks modellər.
- Ansambl metodları - Random Forests və ya Gradient Boosting Machines kimi proqnozlaşdırıcı performansını yaxşılaşdırmaq üçün bir neçə əsas modeli birləşdirən üsullar.

Bu modellər yeni kredit müraciət edənlər üçün kredit riskini dəqiq proqnozlaşdırmaq məqsədi ilə tarixi məlumatlar əsasında öyrədilir.

5. Modelin qiymətləndirilməsi - kredit skorinqi modelinin fəaliyyəti müxtəlif qiymətləndirmə ölçülərindən istifadə etməklə qiymətləndirilir, o cümlədən:
  - Dəqiqlik - düzgün proqnozlaşdırılan nəticələrin nisbəti (məsələn, defolt və qeyri-defolt). Dəqiqlik kredit skorinqi modelləri üçün mühüm qiymətləndirmə göstəricisidir, lakin o, modelin performansının hərtərəfli qiymətləndirilməsini təmin etmək üçün digər ölçülərlə birlikdə şərh edilməlidir, xüsusən də sinif paylamalarının balanssız olduğu və ya yanlış təsnifat xərclərinin asimmetrik olduğu ssenarilərdə.
  - Həssaslıq və spesifiklik - modelin öhdəliklərini yerinə yetirməyən borcalanları (həssaslıq) və öhdəliklərini yerinə yetirməyən borcalanları (spesifiklik) düzgün müəyyən etmək bacarığı.
  - ROC əyrisinin altında sahə (AUC-ROC-Area Under the ROC Curve ) - kredit skorinqində daha yüksək AUC-ROC dəyəri modelin bütün mümkün qərar hədləri üzrə yaxşı və pis kredit risklərini ayırd etməkdə daha yaxşı olduğunu göstərir. 1-ə yaxın AUC-ROC dəyəri güclü ayır-seçkilik gücünü göstərir, yəni model defolt edənləri defolt olmayanlardan effektiv şəkildə ayıra bilir. Digər tərəfdən, 0,5-ə yaxın bir AUC-ROC dəyəri, modelin performansının kredit öhdəliklərini proqnozlaşdırmaqda təsadüfi şansdan yaxşı olmadığını göstərir.

6. Modelin kalibrlənməsi və optimallaşdırılması - qiymətləndirmədən sonra kredit skoringi modeli proqnozlaşdırıcı performansını yaxşılaşdırmaq üçün optimallaşdırmadan keçə bilər. Bu, aşağıdakıları əhatə edə bilər:

- Model parametrlərinin tənzimlənməsi - modelin performansını optimallaşdırmaq üçün tənzimləmə gücü kimi hiperparametrlərin tənzimlənməsi.
- Həddinin tənzimlənməsi - həssaslıq və spesifiklik arasında istənilən tarazlığa nail olmaq üçün qərar hədlərinin dəyişdirilməsi.
- Geribildirim inteqrasiyası - modelin xüsusiyyətlərini və ya alqoritmlərini təkmilləşdirmək üçün domen ekspertləri və ya maraqlı tərəflərdən rəylərin daxil edilməsi.

Məqsəd kreditörün risk tolerantlığına və biznes məqsədlərinə uyğunlaşdırmaq üçün modeli dəqiq tənzimləməkdir.

7. Qiymətləndirmə və monitorinq - kalibrovka edildikdən sonra təsdiqlənmiş kredit skoringi modeli real dünyada istifadə üçün istehsal mühitində yerləşdirilir. Davamlı monitorinq və təsdiqləmə modelin davamlı dəqiqliyini və effektivliyini təmin etmək üçün vacibdir. Bu daxildir:

- Müntəzəm qiymətləndirmə yoxlamaları - proqnoz gücündə hər hansı sürüşmə və ya pisləşməni aşkar etmək üçün modelin performansını yeni məlumatlara qarşı vaxtaşırı qiymətləndirmək.
- Model yeniləmələri - kredit vermə təcrübələrində, borcalanın davranışında və ya iqtisadi şəraitdə dəyişiklikləri nəzərə almaq üçün modelin lazım olduğu kimi yenilənməsi.

Davamlı yoxlama və monitorinq zamanla modelin aktuallığını və etibarlılığını qorumağa kömək edir. Bu hərtərəfli inkişaf prosesi məlumatlı kreditləşdirmə qərarlarını və effektiv risklərin idarə edilməsi təcrübələrini dəstəkləyən möhkəm və effektiv kredit skoringi modelinin yaradılmasını təmin edir [Bessis Joel, 2002].



## III FƏSİL. PYTHON PROQRAMLAŞDIRMA DİLİNDƏ KREDİT RİSK MODELİNİN QURULMASI

### 3.1. Kredit risk modeli məlumatlarının təmizlənməsi və hazırlanması

Kredit skoring modelləri, bank sektorunda istifadə olunan önəmli modellərdən biridir və məqsədə uyğun olaraq müştərilərin kredit öhdəliklərini necə ödəyəcəkləri barədə proqnozlaşdırmaq üçün istifadə olunur. Bu modellər, müştərilərin ödəniş tarixçələri, kredit məlumatları, əmlak, istehlak və digər məlumatlar kimi müxtəlif məlumatları istifadə edərək müştərilərin kredit riskini qiymətləndirmək üçün inkişaf etmişdir. Bu modelin qurulması prosesi aşağıdakı kimidir:

1. Məlumatların toplanması - bankın mövcud müştəri məlumat bazası, kredit büroları, müştəri əməliyyatları və digər əlaqəli mənbələrdən məlumatlar toplanır. Müxtəlif mənbələrdən toplanan məlumatlar bir araya gətirilir:

```
df= pd.read_csv('Downloads/loan_data.csv')
```

```
pd.options.display.max_columns = None
```

```
df.head()
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership
0	60516983	64537751	20000	20000	20000	36 months	12.29	667.06	C	C1	Accounting Clerk	1 year	OWN
1	60187139	64163931	11000	11000	11000	36 months	12.69	369.00	C	C2	Accounts Payable Lead	7 years	MORTGAGE
2	60356453	64333218	7000	7000	7000	36 months	9.99	225.84	B	B3	Nurse	6 years	MORTGAGE
3	59955769	63900496	10000	10000	10000	36 months	10.99	327.34	B	B4	Service Manager	10+ years	MORTGAGE
4	58703693	62544456	9550	9550	9550	36 months	19.99	354.87	E	E4	NaN	NaN	RENT

Şək. 3.1. Kredit skoring data

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 421094 entries, 0 to 421093

Data columns (total 74 columns):

2. Hədəf dəyişəninin müəyyənləşdirilməsi - hədəf dəyişəninin müəyyən edilməsi modelimizi qurmaq və nəticələrimizi qiymətləndirmək üçün çox vacibdir. Modelimizi qurmaq üçün biz hədəf dəyişənini proqnozlaşdırmağa və modelimizin düzgünlüyünü qiymətləndirmək üçün onu faktiki dəyərlərlə müqayisə etməyə çalışırıq. Bunun üçün 'loan\_status' sütununu götürürük, unikal dəyərlərin sayını və faizini tapırıq. Daha sonra bu sütuna istinadən başqa bir sütun (good\_bad) yaradıırıq, lazımi şərtlər verilir və sonda 'loan\_status' sütunu silinir:

```
loan_data['good_bad'] = np.where(loan_data.loc[:,
'loan_status'].isin(['Charged Off', 'Default', 'Late (31-120 days)', 'Charged
Off']), 0, 1)

loan_data.drop(columns = ['loan_status'], inplace = True)
```

3. Məlumatların bölünməsi - məlumat dəstini təlim və sınaq dəstlərinə bölərkən pis kredit tətbiqlərinin nisbətini qoruyur. Daha sonra o, bizə məlumatları manipulyasiya etməyə imkan verən təlim və test dəstlərinin nüsxəsini yaradır:

```
X = loan_data.drop('good_bad', axis = 1)

y = loan_data['good_bad']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2,
random_state = 42, stratify = y)

X_train, X_test = X_train.copy(), X_test.copy()
```

4. Məlumat təmizlənməsi – təmizlənmiş bir verilənlər dəsti, bizə daha doğru nəticələr əldə etməyə və daha etibarlı qərarlar verməyə imkan verir. Bunun üçün biz datamızdakı lazımsız hissələri silirik:

```
na_values = loan_data.isnull().mean()
```

```
na_values[na_values>0.8]
```

```
loan_data.dropna(thresh = loan_data.shape[0]*0.2, how = 'all', axis = 1, inplace
= True)
```

```
loan_data.drop(columns = ['id', 'member_id', 'sub_grade', 'emp_title', 'url', 'title',
'zip_code', 'next_pymnt_d', 'recoveries', 'collection_recovery_fee',
'total_rec_prncp', 'total_rec_late_fee'], inplace = True)
```

```
def emp_length_converter(df, column):
```

```
df[column] = df[column].str.replace('\+ years', '')
```

```
df[column] = df[column].str.replace('< 1 year', str(0))
```

```
df[column] = df[column].str.replace(' years', '')
```

```
df[column] = df[column].str.replace(' year', '')
```

```
df[column] = pd.to_numeric(df[column])
```

```
df[column].fillna(value = 0, inplace = True)
```

```
emp_length_converter(X_train, 'emp_length')
```

```
X_train['emp_length'].unique()
```

```
def loan_term_converter(df, column):
```

```
df[column] = pd.to_numeric(df[column].str.replace(' months', ''))
```

```
loan_term_converter(X_train, 'term')
```

Datmamızdakı tarix sütunlarını götürüb datetime formatına çevirib, cari tarix ilə müvafiq tarix arasındakı fərqi yeni bir sütun olaraq yaradıırıq və X\_train üzərinə tətbiq edilir və köhnə tarix sütunları silinir ('earliest\_cr\_line', 'issue\_d', 'last\_pymnt\_d', 'last\_credit\_pull\_d'):

```
def date_columns(df, column):

    today_date = pd.to_datetime('2024-05-01')

    df[column] = pd.to_datetime(df[column], format = "%b-%y")

    df['mths_since_' + column] = round(pd.to_numeric(((today_date - df[column]) /
    np.timedelta64(1, 'M'))))

    df['mths_since_' + column] = df['mths_since_' + column].apply(lambda x:
    df['mths_since_' + column].max() if x < 0 else x)

    df.drop(columns = [column], inplace = True)

date_columns(X_train, 'earliest_cr_line')

date_columns(X_train, 'issue_d')

date_columns(X_train, 'last_pymnt_d')

date_columns(X_train, 'last_credit_pull_d')

print(X_train['mths_since_earliest_cr_line'].describe())

print(X_train['mths_since_issue_d'].describe())

print(X_train['mths_since_last_pymnt_d'].describe())

print(X_train['mths_since_last_credit_pull_d'].describe())
```

5. Xüsusiyyət seçimi (Feature selection) - verilənlər toplusunda ən vacib xüsusiyyətləri müəyyən edərək maşın öyrənmə modelinin işini yaxşılaşdırmaq

üçün lazımsız və ya daha az məlumat verən xüsusiyyətlərin silinməsidir. Bunun üçün əvvəlcə təlim məlumatlarını kateqoriyalı və ədədi alt qruplara bölək, Chi-squared istifadə edərək kateqoriyalı dəyişənlərə tətbiq edək. Hədəf dəyişəni ilə chi-squared-i hesablamaq üçün təlim dəstindəki hər bir sütun üzərində döngə (loop) verək:

```
X_train_cat = X_train.select_dtypes(include = 'object').copy()

X_train_num = X_train.select_dtypes(include = 'number').copy()

chi2_check = {}

for column in X_train_cat:

    chi, p, dof, ex = chi2_contingency(pd.crosstab(y_train, X_train_cat[column]))

    chi2_check.setdefault('Feature', []).append(column)

    chi2_check.setdefault('p-value', []).append(round(p, 10))

chi2_result = pd.DataFrame(data = chi2_check)

chi2_result.sort_values(by = ['p-value'], ascending = True, ignore_index =

True, inplace = True

chi2_result
```

İndi isə ədədi xüsusiyyətlərin hədəf dəyişəni ilə əlaqəsini qiymətləndirək və F-statistikası və p-dəyərləri tətbiq edək. Daha yüksək F statistik dəyərlər xüsusiyyətin hədəf dəyişənlə daha güclü əlaqəyə malik olduğunu göstərir:

```
X_train_num.fillna(X_train_num.mean(), inplace = True)

F_statistic, p_values = f_classif(X_train_num, y_train)

F_table = pd.DataFrame(data = {'Numerical_Feature':

X_train_num.columns.values, 'F-Score': F_statistic, 'p values':
```

```
p_values.round(decimals=10))
```

```
F_table.sort_values(by = ['F-Score'], ascending = False,
```

```
ignore_index = True, inplace = True)
```

```
F_table
```

6. Multikollinearlığın aşkar olunması – bunun üçün ən yaxşı seçilmiş ədədi xüsusiyyətlər arasındakı əlaqəni vizuallaşdırmaq:

```
top_num_features = F_table.iloc[:33,0].to_list()
```

```
corrmat = X_train_num[top_num_features].corr()
```

```
mask = np.triu(np.ones_like(corrmat, dtype=np.bool))
```

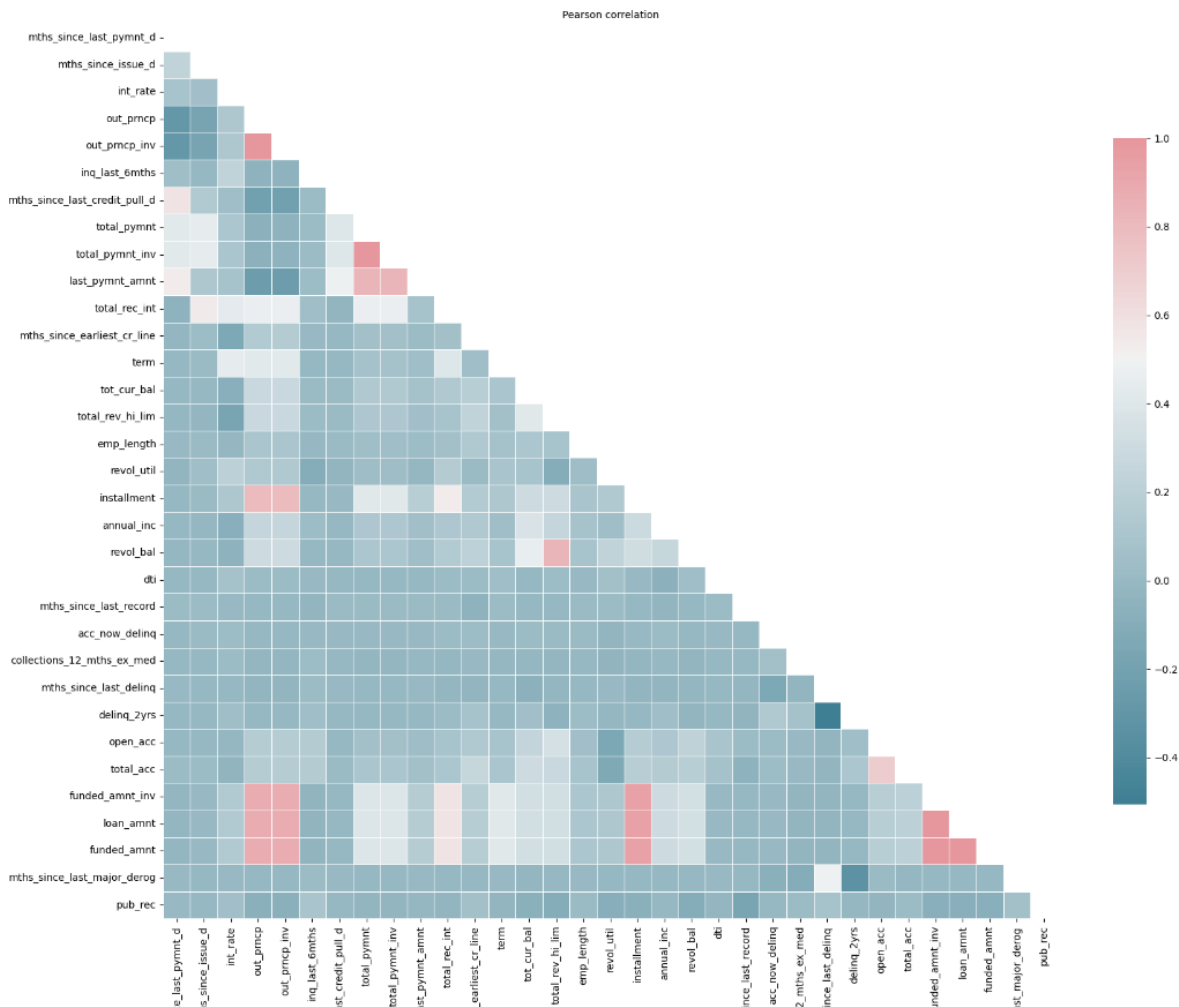
```
f, ax = plt.subplots(figsize=(20,20))
```

```
cmap = sns.diverging_palette(220, 10, as_cmap=True)
```

```
sns.heatmap(corrmat, cmap= cmap, mask= mask, vmax=1, center=0.5,
```

```
square=True, linewidths=.5, cbar_kws={"shrink": .6},annot=True)
```

```
plt.title("Pearson correlation", fontsize =10)
```



Şək. 3.2. Ən yaxşı 33 ədədi xüsusiyyətlərin korrelyasiya qrafiki

7. Dummy dəyişənlərin yaradılması - diskret dəyişənləri dummy dəyişənlərə çevirək:

```
def dummy_creation(df, columns_list):
```

```
    df_dummies = [ ]
```

```
    for col in columns_list:
```

```
        df_dummies.append(pd.get_dummies(df[col], prefix = col, prefix_sep = ':'))
```

```
    df_dummies = pd.concat(df_dummies, axis = 1)
```

```
    df = pd.concat([df, df_dummies], axis = 1)
```

```
return df
```

```
X_train = dummy_creation(X_train, ['grade', 'home_ownership',
'verification_status', 'purpose'])
```

8. Test məlumat dəstini indiyə qədər tətbiq etdiyimiz bütün məlumatların təmizlənməsi prosedurları ilə yeniləyək:

```
emp_length_converter(X_test, 'emp_length')
```

```
date_columns(X_test, 'earliest_cr_line')
```

```
date_columns(X_test, 'issue_d')
```

```
date_columns(X_test, 'last_pymnt_d')
```

```
date_columns(X_test, 'last_credit_pull_d')
```

```
loan_term_converter(X_test, 'term')
```

```
col_to_drop(X_test, drop_columns_list)
```

```
X_test = dummy_creation(X_test, ['grade', 'home_ownership',
'verification_status', 'purpose'])
```

```
X_test = X_test.reindex(labels=X_train.columns, axis=1, fill_value=0)
```

9. WoE Binning/Feature Engineering - WoE metodunu birbaşa orijinal verilənlər toplusuna tətbiq edilməzdən əvvəl verilənlər toplusunun nüsxələrində sınaqdan keçirilə bilər və dəyişikliklərin nəticələri proqnozlaşdırıla bilər:

```
X_train_prepr = X_train.copy()
```

```
y_train_prepr = y_train.copy()
```

```
X_test_prepr = X_test.copy()
```



```
y_test_prepr = y_test.copy()
```

Diskret xüsusiyyətlərin WoEs və IV-lərini təhlil etmək. Bu zaman aşağıdakı kod kateqoriyalı dəyişən üçün WoE çevrilməsini hesablamaq funksiyasını müəyyən edir. Bu transformasiya hər bir kateqoriya üçün WoE dəyərlərini hesablayır və həmçinin hər bir kateqoriya üçün məlumat dəyərini (information value) hesablayır. Nəticədə, o, WoE transformasiyasını və IV (information value) dəyərini ehtiva edən məlumat dəstini qaytarır:

```
def woe_discrete(df, cat_variabe_name, y_df):

    df = pd.concat([df[cat_variabe_name], y_df], axis = 1)

df=pd.concat([df.groupby(df.columns.values[0],as_index=False)[df.columns.values[
]].count(),df.groupby(df.columns.values[0],as_index=False)[df.columns.values[1]].m
ean()], axis = 1)

df = df.iloc[:, [0, 1, 3]]

df.columns = [df.columns.values[0], 'n_obs', 'prop_good']

df['prop_n_obs'] = df['n_obs'] / df['n_obs'].sum()

df['n_good'] = df['prop_good'] * df['n_obs']

df['n_bad'] = (1 - df['prop_good']) * df['n_obs']

df['prop_n_good'] = df['n_good'] / df['n_good'].sum()

df['prop_n_bad'] = df['n_bad'] / df['n_bad'].sum()

df['WoE'] = np.log(df['prop_n_good'] / df['prop_n_bad'])

df = df.sort_values(['WoE'])

df = df.reset_index(drop = True)

df['diff_prop_good'] = df['prop_good'].diff().abs()
```

```

df['diff_WoE'] = df['WoE'].diff().abs()

df['IV'] = (df['prop_n_good'] - df['prop_n_bad']) * df['WoE']

df['IV'] = df['IV'].sum()

return df

```

10. WoE Binning/Reengineering üçün xüsusi sinif müəyyən edilməsi – biz burada hər hansı bir verilənlər toplusunda bütün binning çevrilmələrinə diqqət etmək üçün xüsusi scikit-learn sinfi yaradacağıq və bu fərdi sinif bizə k qat çarpaz doğrulamasını tətbiq etməyə kömək edəcək. Aşağıdakı kod hər bir xüsusiyyət üçün istinad kateqoriyalarını əhatə edən siyahı yaradır. Bu istinad kateqoriyaları WoE çevrilmələrini həyata keçirərkən müqayisə nöqtələri kimi istifadə olunur. Hər bir kateqoriya dəyişən üçün istinad dəyəri müəyyən edirik və bu dəyərlər modelə lazımi müqayisələr aparmağa imkan verir:

```

ref_categories = ['mths_since_last_credit_pull_d:>75',
'mths_since_issue_d:>122', 'mths_since_earliest_cr_line:>434',
'total_rev_hi_lim:>79,780', 'total_rec_int:>7,260', 'total_pymnt:>25,000',
'out_prncp:>15,437', 'revol_util:>1.0', 'inq_last_6mths:>4', 'dti:>35.191',
'annual_inc:>150K', 'int_rate:>20.281', 'term:60',
'purpose:major_purch__car__home_impr', 'verification_status:Not Verified',
'home_ownership:MORTGAGE', 'grade:G']

```

İndi isə verilənlər dəstini modelləşdirmə üçün hazırlayan, WoE kateqoriyalarını yaradan və xüsusiyyətləri uyğun olaraq ayıran və kodlayan geniş transformasiya prosesini əhatə edək. Bu ön emal mərhələsi kredit riski və təsnifat problemlərində çox

önəmlidir. Ona görə ki, hər bir intervalın nəticəyə təsirini başa düşmək və güclü proqnozlaşdırıcı modellər yaratmaq üçün vacibdir:

```
class WoE_Binning(BaseEstimator, TransformerMixin):
    def __init__(self, X):
        self.X = X
    def fit(self, X, y = None):
        return self
    def transform(self, X):
        X_new = X.loc[:, 'grade:A': 'grade:G']
        X_new['home_ownership:OWN'] = X.loc[:, 'home_ownership:OWN']
        X_new['home_ownership:MORTGAGE'] = X.loc[:, 'home_ownership:MORTGAGE']
        X_new['home_ownership:OTHER_NONE_RENT'] = sum([X['home_ownership:ANY'], X['home_ownership:RENT']])
        X_new = pd.concat([X_new, X.loc[:, 'verification_status:Not Verified': 'verification_status:Verified']], axis = 1)

        X_new['purpose:debt_consolidation'] = X.loc[:, 'purpose:debt_consolidation']
        X_new['purpose:credit_card'] = X.loc[:, 'purpose:credit_card']
        X_new['purpose:major_purch_car_home Impr'] = sum([X['purpose:major_purchase'], X['purpose:car'], X['purpose:home_improvement']])
        X_new['purpose:educ_ren_en_sm_b_mov'] = sum([X['purpose:educational'], X['purpose:renewable_energy'], X['purpose:small_business'],
                                                    X['purpose:moving']])
        X_new['purpose:vacation_house_wedding_med_oth'] = sum([X['purpose:vacation'], X['purpose:house'], X['purpose:wedding'],
                                                            X['purpose:medical'], X['purpose:other']])

-----

X_new['mths_since_last_credit_pull_d:missing'] = np.where(X['mths_since_last_credit_pull_d'].isnull(), 1, 0)
X_new['mths_since_last_credit_pull_d:<56'] = np.where((X['mths_since_last_credit_pull_d'] <= 56), 1, 0)
X_new['mths_since_last_credit_pull_d:56-61'] = np.where((X['mths_since_last_credit_pull_d'] > 56) & (X['mths_since_last_credit_pull_d'] <= 61), 1, 0)
X_new['mths_since_last_credit_pull_d:>75'] = np.where((X['mths_since_last_credit_pull_d'] > 61), 1, 0)
X_new.drop(columns = ref_categories, inplace = True)
return X_new
```

Şək. 3.3. WoE Binning/Reengineering üçün xüsusi sinif müəyyənləşdirilməsi

### 3.2. Kredit risk modelinin inkişafı və qiymətləndirilməsi

1. PD Model (Probability of Default) - bu model adətən kredit riskinin idarə edilməsində istifadə edilən qiymətləndirmə modelidir və borcalanın müəyyən zaman aralığında borcunu ödəyə bilməyəcəyi ehtimalını qiymətləndirmək üçün istifadə olunur. Buarada ilk öncə, təlim və test məlumat dəstlərinin formalarını (sətir və sütun nömrələrini) yoxlayacağıq. Bu bizə hər bir məlumat dəstinin ölçülərini (sətir və sütunların sayını) göstərir və məlumat dəstlərinin gözləndiyi kimi olmasını yoxlamağa kömək edir. Sonra Logistic Regression modelinin performansını qiymətləndirmək üçün AUROC və onun əsasında hesablanmış Gini qiymətini hesablayırıq (Mean AUROC: 0.7972 , Gini: 0.5945). Bu qiymətlər modelin təsnifat performansının nə qədər yaxşı olduğunu və sinif balanssızlığını nəzərə alaraq nə qədər effektiv olduğunu göstərir. Daha sonra xülasə cədvəli

(summary\_table) yaradırıq. Bu cədvəl hər bir xüsusiyyət üçün əmsalları və modelin kəşimə dəyərini özündə ehtiva edir:

	Feature name	Coefficients
0	Intercept	1.089633
1	grade:A	1.315895
2	grade:B	0.668634
3	grade:C	0.387743
4	grade:D	0.693195
...	...	...
78	mths_since_issue_d:89-100	0.000000
79	mths_since_issue_d:100-122	0.000000
80	mths_since_last_credit_pull_d:missing	-3.638116
81	mths_since_last_credit_pull_d:<56	0.000000
82	mths_since_last_credit_pull_d:56-61	0.000000

Şək. 3.4. Logistic Regression Model

2. Proqnoz (Prediction) – bu proses kredit verənlərə və risk idarəçilərinə borcalanların risk səviyyələri haqqında daha aydın məlumatlar verir. Bu məlumatlar, kredit portfelinin idarə olunması, kredit məhsullarının qiymətləndirilməsi və borcalanların risklərinin idarə olunması üçün önəmlidir. Burada biz, test məlumat dəsti üzərində modelin ehtimallarını hesablayırıq və hədəf dəyişənin (borcalanların vəziyyətinin) müəyyən olunması üçün modelin təxmin etdiyi ehtimalı göstəririk və bu ehtimalların sadəcə pozitiv sinif (default - standartdan ayrılmış kateqoriya) üzərindəki ehtimallarını seçirik. Daha sonra faktiki sinifləri və proqnozlaşdırılan ehtimalları ehtiva edən yeni DataFrame yaradırıq. Faktiki y\_test DataFrame indeksini düzgün şəkildə sıfırlamaq üçün y\_test\_temp DataFrame yaradırıq və onun y\_hat\_test\_proba ilə düzgün birləşməsinə imkan veririk. Sonra, nəticənin y\_test ilə eyni sayda sətir olub-olmadığını yoxlamaq üçün forma funksiyasından istifadə edirik:

```
y_test_proba.columns = ['y_test_class_actual', 'y_hat_test_proba']
```

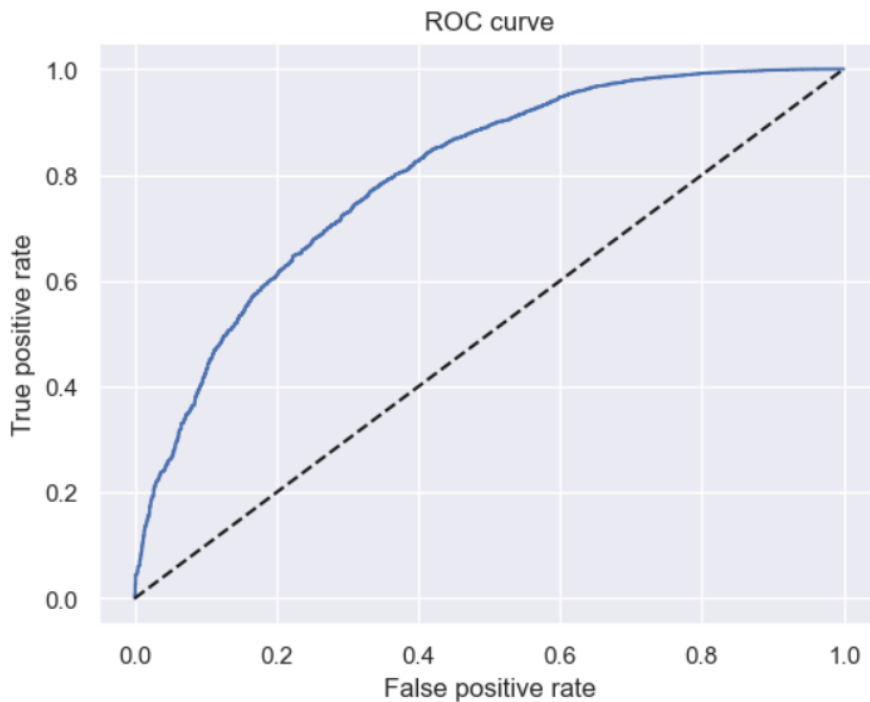
```
y_test_proba.index = X_test.index
```

```
y_test_proba
```

	<b>y_test_class_actual</b>	<b>y_hat_test_proba</b>
<b>85226</b>	1	0.767027
<b>135248</b>	1	0.575147
<b>124342</b>	1	0.706738
<b>318929</b>	1	0.695942
<b>49176</b>	1	0.625467
...	...	...
<b>287341</b>	1	0.946924
<b>401921</b>	1	0.190760
<b>63105</b>	1	0.844285
<b>358896</b>	1	0.335628
<b>118334</b>	1	0.897676

Şək. 3.5. Test verilənlər dəstinin proqnoz və ehtimalları

3. Confusion Matrix & AUROC on test set - test məlumat dəsti üzərində bir Confusion Matrix və AUROC hesablayacağıq. Burada test məlumat dəstindəki faktiki siniflər ilə model tərəfindən proqnozlaşdırılan siniflər arasındakı əlaqəni qiymətləndiririk. İlk öncə bir threshold dəyəri müəyyən edirik, sonra proqnozlaşdırılan ehtimallara əsasən yeni sinif müəyyən edirik, yəni ehtimal həddi qiymətdən böyükdürsə 1 (müsbət), əks halda 0 (mənfi) kimi təsnif edirik. Nəhayət, bu proqnozlar əsasında mürəkkəblik matrisi yaradırıq. Confusion matrix faktiki və proqnozlaşdırılan siniflər arasındakı əlaqəni göstərir. Sonra, faktiki siniflər və proqnozlaşdırılan ehtimallar arasındakı əlaqəni təmsil edən müxtəlif hədd dəyərlərində false positive rate (FPR) və true positive rate (TPR) dəyərlərini hesablayırıq. Sonra x oxuna uyğun FPR və y oxuna uyğun TPR ilə qrafik çəkirik:

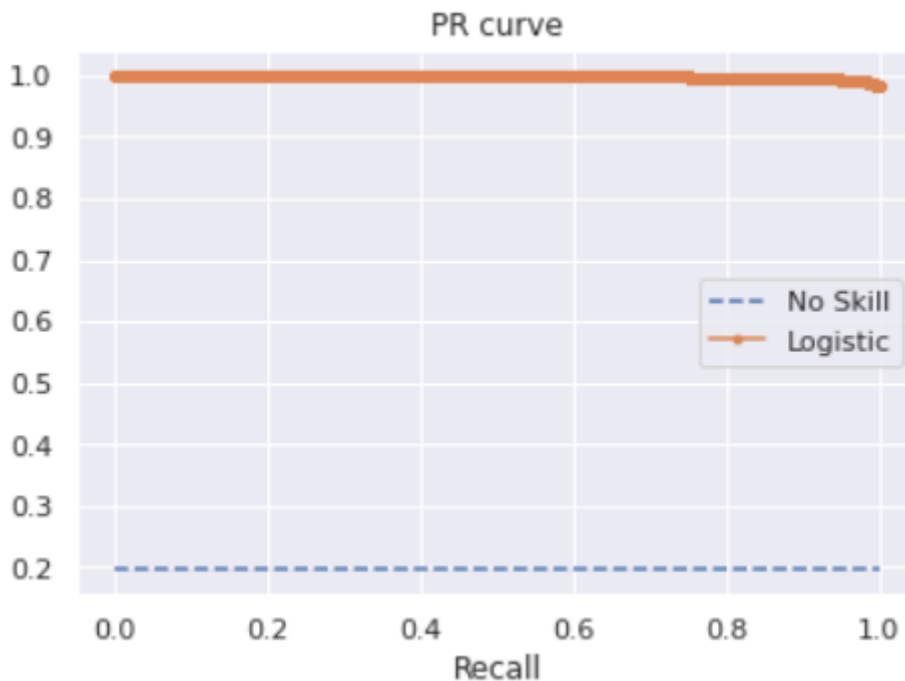


Şək. 3.6. ROC əyrisi

Qrafik hazır olduqdan sonra, ROC əyrisi altındakı sahəni hesablayırıq və ROC əyrisinin dəyərini əldə etikdən sonra, Gini əmsalını hesablayırıq (Gini əmsalı təsnifat modelinin performansını ölçmək üçün istifadə olunan metrikdir), yəni

$$\text{Gini} = \text{AUROC} * 2 - 1 \quad (3.1)$$

Daha sonra precision və recall arasındakı əlaqəni vizuallaşdırmaq üçün Precision-Recall (PR) əyrisini müəyyənləşdiririk. Müsbət sinifin nisbətini göstərən "No Skill" adlı bir xətt də çəkirik və PR əyrisi altındakı sahəni təmsil edən PR AUC dəyərini ( 0.997578707851294) hesablayırıq. Daha yüksək PR AUC dəyəri modelin daha yaxşı performans göstərdiyini bildirir:



Şək. 3.7. Precision-Recall əyrisi

4. Modelin tətbiqi – burada biz 'ref\_categories' siyahısındakı dəyərlərdən ibarət və 'Feature name' adlı sütundan ibarət yeni DataFrame yaradıırıq. Sonra, "Coefficients" adlı başqa bir sütun yaradıırıq və bu sütuna hələki 0 dəyərini yazırıq. Sonra 'summary\_table' və 'df\_ref\_categories' CONCAT edirik, yəni birləşdiririk və 'df\_scorecard' DataFrame-i yaradıırıq. Sonra df\_scorecard DataFrame-ə " Original feature name" adlı yeni sütun əlavə edirik və bu sütun bizə "Feature name" sütunundan lazımi hissələri özündə ehtiva edir:

	index	Feature name	Coefficients	Original feature name
0	0	Intercept	1.089633	Intercept
1	1	grade:A	1.315895	grade
2	2	grade:B	0.668634	grade
3	3	grade:C	0.387743	grade
4	4	grade:D	0.693195	grade
...	...	...	...	...
95	12	term:60	0.000000	term
96	13	purpose:major_purch_car_home_impr	0.000000	purpose
97	14	verification_status:Not Verified	0.000000	verification_status
98	15	home_ownership:MORTGAGE	0.000000	home_ownership
99	16	grade:G	0.000000	grade

Şək. 3.8. Referans kateqoriyalar üzrə scorecard

İndi kredit skoring minimum və maksimum həddi müəyyən edək, yəni `min_score = 300` və `max_score = 850` olsun. Sonra scorecard yaratmaq üçün lazımı hesablamaları edək:

```
min_sum_coef = df_scorecard.groupby('Original feature name')['Coefficients'].min().sum()
```

```
max_sum_coef = df_scorecard.groupby('Original feature name')['Coefficients'].max().sum()
```

```
df_scorecard['Score - Calculation'] = df_scorecard['Coefficients'] * (max_score - min_score) / (max_sum_coef - min_sum_coef)
```

```
df_scorecard.loc[0, 'Score - Calculation'] = ((df_scorecard.loc[0, 'Coefficients'] - min_sum_coef) / (max_sum_coef - min_sum_coef)) * (max_score - min_score) + min_score
```

```
df_scorecard['Score - Preliminary'] = df_scorecard['Score - Calculation'].round()
```

```
df_scorecard
```



Scorecard-in minimum və maksimum mümkün ümumi xallarını yoxlasaq,  $\text{min\_score} = 300$  və  $\text{max\_score} = 851$  alınır. Sonra biz Score – Preliminary və Score – Calculation arasında fərqi tapırıq və bu fərqləri ‘Difference’ adlı sütuna daxil edək və son olaraq ‘Intercept’ üçün dəyəri azaldırıq və 653 daxil edirik:

	index	Feature name	Coefficients	Original feature name	Score - Calculation	Score - Preliminary	Difference	Score - Final
0	0	Intercept	1.089633	Intercept	666.654357	667.0	0.345643	653.0
1	1	grade:A	1.315895	grade	38.815440	39.0	0.184560	39.0
2	2	grade:B	0.668634	grade	19.722944	20.0	0.277056	20.0
3	3	grade:C	0.387743	grade	11.437383	11.0	-0.437383	11.0
-----								
97	14	verification_status:Not Verified	0.000000	verification_status	0.000000	0.0	0.000000	0.0
98	15	home_ownership:MORTGAGE	0.000000	home_ownership	0.000000	0.0	0.000000	0.0
99	16	grade:G	0.000000	grade	0.000000	0.0	0.000000	0.0

### Şək. 3.9. Scorecard hesablamaları

Nəhayət, bizim scorecard üçün mümkün dəyərlərimiz  $\text{min\_score} = 286$  və  $\text{max\_score} = 837$  oldu.

5. Test məlumat dəstindəki bütün müşahidələr üçün kredit ballarının hesablanması - yəni, SBu sinif xüsusiyyətləri düzgün manipulyasiya etmək və kodlaşdırmaq üçün lazımi çevrilmələri həyata keçiririk. Daha sonra modelin gözləntilərinə uyğunlaşmaq üçün dəyişdirilmiş test dəstinə ‘Intercept’ sütununu əlavə edirik. Bu, modelin hər bir müşahidə üçün müvafiq proqnozlar verə bilməsində bizə kömək edəcək. Sonra, test dəstindəki müşahidələr üçün kredit skoring hesablamaq üçün scorecard ballarından istifadə edirik. Nəticədə hər bir müşahidənin kredit skoring müəyyən edirik:

	Score
<b>1645</b>	477.0
<b>2243</b>	482.0
<b>956</b>	488.0
<b>1810</b>	492.0
<b>770</b>	492.0
...	...
<b>364565</b>	747.0
<b>323302</b>	748.0
<b>335833</b>	748.0
<b>341602</b>	749.0
<b>211824</b>	754.0

Şək. 3.10. Skoring cədvəli

İndi isə ROC əyrisindən əldə edilən müxtəlif hədd dəyərlərinə uyğun kredit skoring hesablayırıq və onları DataFrame-də saxlayırıq. Bu məlumat kredit skoringi müəyyən həddlərə uyğunlaşdırmaq və kredit müraciətlərini qiymətləndirərkən qərar qəbul etmə prosesini optimallaşdırmaq üçün istifadə olunur:

```
df_cutoffs = pd.DataFrame(thresholds, columns = ['thresholds'])
```

```
df_cutoffs['Score'] = ((np.log(df_cutoffs['thresholds'] / (1 - df_cutoffs['thresholds'])) - min_sum_coef) * ((max_score - min_score) / (max_sum_coef - min_sum_coef)) + min_score).round()
```

```
df_cutoffs
```

	<b>thresholds</b>	<b>Score</b>
<b>0</b>	0.007206	489.0
<b>1</b>	0.009032	496.0
<b>2</b>	0.010893	502.0
<b>3</b>	0.012200	505.0
<b>4</b>	0.012369	505.0
...	...	...
<b>80339</b>	0.986383	761.0
<b>80340</b>	0.986479	761.0
<b>80341</b>	0.986822	762.0
<b>80342</b>	0.987011	762.0
<b>80343</b>	0.988913	767.0

Şək. 3.11. Thresholds və ballar

Sonra, 'y\_hat\_test\_proba' sütununda təxmin edilən ehtimallar hədd dəyərindən (p) böyük və ya ona bərabər olduqda 1, əks halda 0 qaytaran 'y\_test\_proba' DataFrame-də NumPy massivi yaradırıq və 1 (0) dəyərlərini cəmləyirik, bununla da təsdiq edilmiş kredit müraciətlərinin sayını göstərmiş oluruq:

```
def n_approved(p):
    return np.where(y_test_proba['y_hat_test_proba'] >= p, 1, 0).sum()
df_cutoffs['N Approved'] = df_cutoffs['thresholds'].apply(n_approved)
df_cutoffs['N Rejected'] = y_test_proba['y_hat_test_proba'].shape[0] -
df_cutoffs['N Approved']
df_cutoffs['Rejection Rate'] = 1 - df_cutoffs['Approval Rate']
df_cutoffs
```

	thresholds	Score	N Approved	N Rejected	Approval Rate	Rejection Rate
<b>0</b>	0.007206	489.0	84219	0	1.000000	0.000000
<b>1</b>	0.009032	496.0	84218	1	0.999988	0.000012
<b>2</b>	0.010893	502.0	84217	2	0.999976	0.000024
<b>3</b>	0.012200	505.0	84216	3	0.999964	0.000036
<b>4</b>	0.012369	505.0	84215	4	0.999953	0.000047
...	...	...	...	...	...	...
<b>80339</b>	0.986383	761.0	5	84214	0.000059	0.999941
<b>80340</b>	0.986479	761.0	4	84215	0.000047	0.999953
<b>80341</b>	0.986822	762.0	3	84216	0.000036	0.999964
<b>80342</b>	0.987011	762.0	2	84217	0.000024	0.999976
<b>80343</b>	0.988913	767.0	1	84218	0.000012	0.999988

Şək. 3.12. Thresholds və təsdiq statistikasını

```
df_cutoffs[df_cutoffs['thresholds'].between(0.24448, 0.24489)]
```

	thresholds	Score	N Approved	N Rejected	Approval Rate	Rejection Rate
<b>4186</b>	0.244488	601.0	79975	4244	0.949608	0.050392
<b>4187</b>	0.244527	601.0	79974	4245	0.949596	0.050404
<b>4188</b>	0.244560	601.0	79973	4246	0.949584	0.050416
<b>4189</b>	0.244584	601.0	79972	4247	0.949572	0.050428
...	...	...	...	...	...	...
<b>4201</b>	0.244772	601.0	79960	4259	0.949429	0.050571
<b>4202</b>	0.244827	601.0	79959	4260	0.949418	0.050582
<b>4203</b>	0.244887	601.0	79958	4261	0.949406	0.050594

Şək. 3.13. Xüsusi diapazonda thresholds və təsdiq statistikasını

Bu məlumatlar kredit müraciətlərini qiymətləndirmək üçün istifadə edilən həddin hansı təsirə malik olduğunu anlamaq üçün istifadə olunur. Optimal hədd diapazonuna düşən bu məlumatlar kreditin təsdiqlənməsi və rədd edilməsi dərəcələrinin müəyyən edilməsində əhəmiyyətli rol oynayır və kredit siyasətlərinin optimallaşdırılmasında bizə kömək edir.

Nəticə etibarlı ilə kredit riskinin modelləşdirilməsi bankın gələcəkdə hansı müştərilərin kreditləri üzrə defolt edə biləcəyini proqnozlaşdırmaq qabiliyyətini təkmilləşdirmək məqsədi daşıyır. Biz kredit-defoltu təhlil edərək və gələcəkdə hansı kreditlərin defolt ehtimalını müəyyən etməklə, bank ümumi riskini azaltmaq və gəlirliliyini artırmaq üçün müvafiq tədbirlər görə bilər. Statistik üsullardan və maşın öyrənmə alqoritmlərindən istifadə etməklə biz böyük həcmdə kredit məlumatlarını təhlil edə və defolt ehtimalının daha yüksək olduğunu göstərən nümunələri və meylləri müəyyən edə bildik.

## NƏTİCƏ

Disertasiya işi bank sektorunda Data Mining metodlarının tətbiqi ilə məlumatların effektiv şəkildə analiz edilməsi və işlənməsi məsələlərinə həsr edilmişdir. Bu tətbiqlər şirkətlərin məlumatları daha ətraflı şəkildə emal edilməsi, proqnozların verilməsi, müştərilərin davranışlarını müəyyən edilməsinə və iş proseslərini təkmilləşdirilməsinə kömək edir.

Bank işinin planlaşdırılması gələcək üçün hədəflərin müəyyən edilməsi və onlara nail olmaq yollarının işlənilib hazırlanması prosesidir.

Risqlərin idarə edilməsi bank fəaliyyətinin hərtərəfli və kompleks qiymətləndirilməsini - güclü və zəif tərəflər, təşkilati və bankın strukturu, maliyyə nəticələri, kadr siyasəti, nəzarət funksiyalarını tələb edir.

Əsas məqsəd, bank sahəsində Big Data və Data Mining texnologiyalarının tətbiqi ilə bank sektorunun effektiv idarə ediməsidir. Dissertasiya işində banklarda müştəri tələblərinin daha yaxşı anlaşılması, maliyyə risklərinin daha effektiv idarə olunması, yeni maliyyə məhsullarının inkişafı və bankların strateji qərarlarında biliyə əsaslanan məlumatların istifadə edilməsinin təsirini qiymətləndirilməsi məsələlərinə baxılır.

Nəticə olaraq, bu disertasiya işi, bank sahəsində Big Data və Data Mining texnologiyalarının tətbiqi ilə bankların maliyyə sektorunda effektiv məlumat idarəetməsinin və inkişafının artırılması üçün vacib olan biliyə əsaslanan qərarların qəbul edilməsinə kömək edir.

## ƏDƏBİYYAT

1. Barske A.E. (2013-2014). *Big Data Business Guide*.
2. Bessis J.J. (2002). *Risk Management in Banking*.
3. Bramer M.A. (2007). *Principles of Data Mining*.
4. Charu C. A. (2015). *Data Mining: The Textbook*.
5. Chen H., Chiang R.H., Storey V.C. (2012). *Business intelligence and analytics: from big data to big impact. Journal MIS Quarterly* ( Vol. 36, pp. 1165-1188).
6. Clifford L.S., (2008). *Big data: How do your data grow?*
7. Daniel T. L., Chantal D. L. (2023). *Data Mining and Predictive Analytics*.
8. Əliquliyev R.M., Hacırahimova M. Ş. (2014). "Big Data" phenomenon: problemlər və imkanlar. *İnformasiya texnologiyaları problemləri* ( № 2, səh. 3-16) .
9. Fan J.U., Han F.A., Liu H.H. (2014). Challenges of Big Data analysis. *National Science Review* (Vol. 1, pp. 2-8).
10. Flach P.A. (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*.
11. Gandomi A.H., Haider M.A. (2015). Beyond the hype: Big data concepts, methods and analytics. *TedInternational Journal of Information Management* ( Vol . 35, pp. 137–144).
12. Gorodov E.A., Gubarev V.G. (2013). Analytical Review of Data Visualization Methods in Application to Big Data. *Journal of Electrical and Computer Engineering* (Vol. 3, pp.1-7).
13. Gregory E.B., Akhgar B.B., (2015). *Application of Big Data for National Security, Butterworth-Heinemann*.
14. Han J.J., Kamber M.K., Pei J.D. (2021). *Data Mining: Concepts and Techniques*.
15. Hand D.D., Mannila H.O., Smyth P.S. (2001). *Principles of Data Mining*.
16. Husain M.S., Khan M.Z., (2020). *Big Data: Concepts, Methodologies, Tools and Applications*.

17. Jagadish H.V. (2015). Big Data and Science: Myths and Reality. *Big Data Research* (Vol. 2, pp. 49–52).
18. Jeroen J.S. (2014). *Data Science at the Command Line*.
19. Jina X.S., Benjamin W.E., Chenga X.A., Wang Y.T. (2015). Significance and Challenges of Big Data Research. *Big Data Research* (Vol. 2, pp. 59–64).
20. Johann T. H., Barbara K.R. (2020). *Big Data in Context: Legal, Social and Technological Insights*.
21. Kambatla K.K., Kollias G.U., Kumar V.H., Grama A.A. (2014). Trends big data analytics. *Parallel and Distributed Computing* (Vol. 74, pp. 2561-2573).
22. Kundu A.U. (2013). *Data Mining: An Overview from Database Perspective*
23. Labrinidis A., J. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment* (Vol. 5, pp. 1-17).
24. Leskovec J.U., Rajaraman A.R., Jeffrey D.U. (2014). *Mining of Massive Datasets* (pp 1-18)..
25. Linoff G.S., Berry M.J., (2021). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* .
26. Madden S.J., (2012). From Databases to Big Data. *IEEE Internet Computing* (Vol. 16, p. 4-6).
27. Maier M.P. (2013). *Towards a Big Data Reference Architecture* .
28. Manyika J.J., Chui M.M., Brown B.H. (2011). *Big data: The next frontier for innovation, competition, and productivity*.
29. Marr W.B., (2015). *Using Smart Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance*.
30. Mohammed J.Z., Wagner M.J. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms* ( pp 370-641).
31. Moore A.W. (2013). *Practical Data Mining Tutorial*.
32. Provost F. A., Fawcett T.E. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*.
33. Pujari A.K. , (2001). *Data Mining Techniques*.



34. Rajkumar B.A., Wang C.F., Sushmita R.A. (2020). *Big Data: Principles and Paradigms*.
35. Sang C. S., Edward J. W. (2021). *Big Data and Visual Analytics*.
36. Tan P.N., Stainbech M.J., Kumar V.K. (2021). *Introduction to Data Mining*.
37. Tene O.P., Polonetsky J.A. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property* (Vol. 11, Issue 5)..
38. Ularu E.G., Puican F.C., Apostu A.A., Velicanu M.S. (2012). Perspectives on Big Data and Big Data Analytics . *Database Systems Journal* ( Vol. 3, pp. 3-15).
39. Walker R.J. (2015). *From big data to big profits*.
40. Witten H.L., Frank E.E., (2020). *Data Mining: Practical Machine Learning Tools and Techniques*.

## S U M M A R Y

The dissertation is dedicated to the effective analysis and utilization of information in the banking sector through the application of Data Mining methods. These applications assist in more thorough utilization of company data, making forecasts, determining customer behaviors, and improving business processes.

The planning of banking operations involves identifying goals for the future and developing strategies to achieve them.

Risk management in banking requires comprehensive and multifaceted assessment of the bank's activities, including strengths and weaknesses, organizational and bank structure, financial results, personnel policies, and control functions.

The main objective is the effective management of the banking sector through the application of Big Data and Data Mining technologies. The dissertation focuses on better understanding customer demands in banks, more effective management of financial risks, development of new financial products, and evaluating the impact of using knowledge-based information in banking strategic decisions.

In conclusion, this dissertation contributes to enhancing the effectiveness of information management and development in the financial sector of banks through the application of Big Data and Data Mining technologies, thereby facilitating the adoption of knowledge-based decisions.

## Р Е З Ю М Е

Диссертация посвящена проблемам эффективного анализа и использования информации в банковском секторе с применением методов Data Mining. Эти приложения помогают более полно использовать данные компаний, делать прогнозы, определять поведение клиентов и улучшать бизнес-процессы.

Планирование банковской деятельности включает в себя определение целей на будущее и разработку стратегий их достижения.

Управление рисками в банковской деятельности требует всесторонней и комплексной оценки - как сильных, так и слабых сторон, организационной структуры банка, финансовых результатов, кадровой политики и функций контроля.

Основная цель - эффективное управление банковским сектором с применением технологий Big Data и Data Mining. Диссертация сосредоточена на лучшем понимании потребностей клиентов в банках, более эффективном управлении финансовыми рисками, развитии новых финансовых продуктов и оценке влияния использования информации, основанной на знаниях, на стратегические решения банков.

В результате, данная диссертация способствует повышению эффективности управления информацией и развитию финансового сектора банков через применение технологий Big Data и Data Mining, что облегчает принятие решений, основанных на знаниях.