

AZƏRBAYCAN RESPUBLİKASI ELM VƏ TƏHSİL NAZİRLİYİ

AZƏRBAYCAN TEXNİKİ UNİVERSİTETİ

YÜKSƏK TƏHSİL İNSTİTUTU

əlyazma hüququnda

Məmmədova Leyla Rəhim qızının

“Maşın öyrənməsindən istifadə etməklə tələbə müvəffəqiyyətini
proqnozlaşdırın modelin qurulması” mövzusunda

MAGİSTRİK DİSSERTASIYASI

060509 Kompüter Elmləri

Kompüterli modelləşdirmə

Elmi rəhbər: Kərim Tahiroğlu

Bakı-2024



AZƏRBAYCAN TEXNİKİ UNİVERESİTETİ
YÜKSƏK TƏHSİL İNSTİTUTU

MAGİSTRANTIN ANDI

Maşın öyrənməsindən istifadə edərək tələbə müvəffəqiyyətini proqnozlaşdıran modelin qurulması mövzusunda təqdim etdiyim magistrlik dissertasiyasını elmi əxlaq normalarına və istinad qaydalarına tam riayət etməklə və istifadə etdiyim bütün mənbələri ədəbiyyat siyahısında əks etdirməklə yazdığımı and içirəm və magistrlik dissertasiyasının AzTU Kitabxana İnformasiya Mərkəzində saxlanması, həmin mərkəz tərəfindən AzTU Rəqəmsal Repozitasiyasına daxil edilərək repozitariyanın veb saytında yerləşdirilməsinə icazə verirəm.

Leyla Məmmədova

Tarix:

MÜNDƏRİCAT

GİRİŞ.....	4
I FƏSİL. TƏLƏBƏ MÜVƏFFƏQİYYƏTİNƏ TƏSİR EDƏN AMİLLƏR.....	7
1.1 Tələbə müvəffəqiyyətinə təsir edən amillər hansılardır?.....	7
1.2 Aşağı performans göstərən tələbələrin üzləşdiyi problemlər.....	11
1.3 Tələbə müvəffəqiyyətinin proqnozlaşdırılmasının təhsildə önəmi.....	12
1.4 Maşın öyrənməsinin təhsil sahəsində tətbiqi.....	14
1.5 Təhsildə maşın öyrənməsinin tətbiqinin faydaları və çətinlikləri.....	16
II FƏSİL MAŞIN ÖYRƏNMƏSİ NƏDİR?.....	19
2.1 Maşın öyrənməsi haqqında məlumat.....	19
2.2 Maşın öyrənməsinin tətbiq sahələri.....	20
2.3 Maşın öyrənməsinin təsnifatı.....	23
2.4 Maşın öyrənmə alqoritmləri.....	32
2.5 Maşın öyrənməsi həyat dövrü.....	47
2.6 Maşın öyrənməsində istifadə olunan vasitələr.....	50
III FƏSİL MODELİN QURULMASI MƏRHƏLƏLƏRİ.....	53
3.1 Datanın toplanılması və vizuallaşdırılması.....	53
3.2 Məlumatların emalı , təmizlənməsi.....	58
3.3 Xüsusiyyət mühəndisliyi.....	61
3.4 Verilənlərin train və test datalarına ayrılması.....	61
3.5 Modelin öyrədilməsi.....	62
3.6 Modelin qiymətləndirilməsi.....	64
NƏTİCƏ.....	71
İSTİFADƏ EDİLMİŞ ƏDƏBİYYAT SİYAHISI.....	72

Mövzunun aktuallığı. Ali təhsil müəssisələrinin əsas məqsədi tələbələrinə keyfiyyətli təhsil verməkdir və bu səbəbdən təhsil səviyyəsinin yüksəldilməsi üçün mövcud problemlər araşdırılmalı və həll edilməlidir. Əksər ali təhsil müəssisələrinin qarşılaşdıqları əsas problemlərdən biri isə zəif akademik nəticələrə sahib tələbələrin olmasıdır. Tələbələrin aşağı akademik göstəricilərə sahib olması onların universiteti vaxtında bitirə bilməməsi, gələcək təhsil və məşğulluq imkanlarının məhdudlaşması, təqaüdlərinin itirilməsi, təhsil müəssisəsinin nüfuzuna xələl gətirməsi və s. kimi problemlərə səbəb ola bilər. Aşağı akademik göstəricilərə sahib tələbələrin üzləşdiyi əsas problemlərdən biri də təhsil xərclərinin artmasıdır. Təhsil xərclərinə təsir edən amillərdən ən əsası isə tələbələrin məzun olmaq üçün oxumağa sərf etdikləri vaxtdır. Məsələn, amerikalı tələbələrin kredit borcu bir çox tələbələrin vaxtında bitirməməsi səbəbindən artmışdır. İraqda ali təhsil dövlət tərəfindən tələbələrə pulsuz verilir. Ancaq vaxtında məzun olmamaq dövlətə əlavə xərclər gətirir. Bu xərclərdən qaçmaq üçün dövlət tələbənin vaxtında məzun olmasını təmin etməlidir. Bu kimi problemlərin qarşısını almaq üçün zəif akademik göstəricilər əldə edən tələbələri daha erkən aşkarlamaq və onlara lazımi dəstəyin göstərilməsi vacibdir. Bu məsələnin həlli üçün tələbələrin akademik müvəffəqiyyətini proqnozlaşdırmaq böyük əhəmiyyət kəsb edir. Tələbə performansını proqnozlaşdırmaq təhsildə bir çox aspektlərdə önəmlidir. Çünki bu yalnız zəif nəticələr göstərən tələbələrin erkən müəyyən edilməsi üçün deyil, eyni zamanda bir çox məsələlərə müsbət təsir göstərir. Həmin məsələlərə tələbə inkişafının monitorinqi, təhsil siyasətinin inkişafı, tələbələrə dəstək xidmətləri, tədris proseslərinin təkmilləşdirilməsi, gələcək uğurun proqnozlaşdırılması, resurslardan səmərəli istifadə, fərdiləşdirilmiş öyrənmə yanaşmaları, tələbə motivasiyasının artırılmasını, təhsildə məlumatlara əsaslanan qərarların qəbulu kimi məsələləri misal göstərmək olar. Bu səbəblərə görə tələbə müvəffəqiyyətinin proqnozlaşdırılması təhsil sisteminin effektivliyi və tələbə uğuru üçün kritik əhəmiyyət kəsb edir və bu məqamlar eyni zamanda təhsil sistemində bu mövzunun müxtəlif aspektlərdə vacibliyini daha da aydın izah edir.

Tədqiqatın məqsədi və vəzifələri. Dissertasiya işinin əsas məqsədi tələbə müvəffəqiyyəni proqnozlaşdıran modelin qurulmasıdır. Model maşın öyrənməsindən istifadə edilərək qurulur. Modelin qurulmasında əsas vəzifələr aşağıdakılardır:

- Verilənlərin toplanılması
- Toplanmış verilənlərin təmizlənməsi
- Verilənlərin train və test datalarına bölünməsi
- Model üçün məsələnin həllinə uyğun olan maşın öyrənmə alqoritmlərinin müəyyən olunması
- Modelin seçilmiş alqoritmlər vasitəsilə öyrədilməsi
- Model vasitəsilə proqnozların edilməsi, tətbiqi
- Modelin qiymətləndirilməsi, ən yaxşı göstəriciyə sahib alqoritmin seçilməsi

Tədqiqatın predmeti və obyektı. Tədqiqatın predmeti yəni ümumi mövzusu tələbə müvəffəqiyyətinin proqnozlaşdırılması məsələsidir. Bu tələbələrin akademik göstəricilərinin əvvəlcədən müəyyən edilməsi prosesini əhatə edir. Tədqiqat obyektı yəni tədqiqatın diqqət mərkəzində olduğu xüsusi başqa sözlə, təfərrüatlı şəkildə araşdırılan və təhlil edilən tədqiqatın elementi isə maşın öyrənməsi vasitəsilə modelin qurulmasıdır.

Tədqiqat metodları. Dissertasiya işində maşın öyrənməsi alqoritmləri vasitəsilə model qurulur. Qurulan model tələbə müvəffəqiyyətinə proqnozlaşdıracağı üçün əsasən maşın öyrənməsinin xətti reqressiya, qərar ağacı, təsadüfi meşə, dəstək vektor reqressiya, Lasso, AdaBoosting reqressor, K-qonşu reqressor kimi məşhur reqressiya alqoritmlərindən istifadə edilmişdir.

Elmi yeniliyin elementləri. Tələbə müvəffəqiyyətinə proqnozlaşdıran model təhsil sistemindəki boşluqların aşkarlanmasına sistemli yanaşmanı təmin edir.

Praktiki həll. Dissertasiya işinin praktiki əhəmiyyəti ondan ibarətdir ki, qurulmuş model vasitəsilə aşağı akademik nəticələr almaq ehtimalı olan tələbələr daha erkən aşkarlanacaq, bu isə öz növbəsində həmin tələbələrə lazımı dəstək verilməsinə beləliklə bu halların qarşısını almaqla akademik göstəriciləri zəif olan tələbələrin

sayının azaldılmasına və müəllimlərin bu məlumatlara əsaslanaraq daha doğru qərarlar verməsinə səbəb olacaq.

Müdafiə üçün təqdim edilən nəticələr. Modelin tətbiqi nəticəsində tələbələrin ümumi orta müvəffəqiyyət göstəriciləri proqnozlaşdırılmışdır və proqnozlaşdırılan qiymətlər əsasında aşağı akademik nəticələr göstərə biləcək tələbələr müəyyən olunmuşdur.

TƏSİR EDƏN AMİLLƏR

1.1 Tələbə müvəffəqiyyətinə təsir edən amillər hansılardır?

Son zamanlar maşın öyrənməsindən yüksək performans və effektiv nəticələr əldə etmək üçün səmərəli şəkildə istifadə edilir. Maşın öyrənmə alqoritmləri vasitəsilə əldə edilən nəticələr əsasında bir çox problemlər proqnozlaşdırıla və həll edilə bilər. Buna görə də müxtəlif sektorlarda maşın öyrənmə alqoritmlərinin tətbiqi son illərdə daha çox məşhurlaşmış. Bu tətbiq sahələrindən biri də təhsil sektorudur. Təhsil sahəsində maşın öyrənməsinin tətbiqi ilə böyük verilənlər bazasını doğru və qısa zaman ərzində analiz etmək, verilənlər əsasında qanunauyğunluqları müəyyən edərək nəticələr əldə etmək və onlar vasitəsilə bir çox təhsildə olan problemləri həll etmək olar. Həll edilməli məsələlərdən biri də tələbə performansının proqnozlaşdırılmasıdır. Beləki, tələbə müvəffəqiyyətinin proqnozlaşdırılması zəif nəticə göstərən tələbələrin erkən müəyyən edilməsi ilə bəzi əsas problemləri həll etməyə kömək edir.

Maşın öyrənməsi tələbə performansını proqnozlaşdırmaq və tələbələri daha effektiv dəstəkləmək üçün güclü bir vasitə kimi istifadə edilə bilər. Maşın öyrənməsi nümunələri tanımaq və onlardan proqnozlar vermək üçün istifadə etmək üçün məlumatları təhlil edir. Maşın öyrənmə alqoritmlərindən istifadə edilərək qurulan model verilmiş məlumat dəstindəki nümunələri tanıyaraq və proqnozlar vermək üçün nümunələrdən istifadə edərək gələcək tələbə akademik uğurunu proqnozlaşdırmaq məqsədi daşıyır. Model qurulmamışdan əvvəl aşağıda göstərilən tədqiqat məqsədləri araşdırılır:

1. Tələbələrin akademik uğurlarına hansı amillər əhəmiyyətli təsir göstərir?

2. Bu amillər vasitəsilə maşın öyrənməsindən istifadə edərək tələbələrin akademik performansını necə proqnozlaşdırıla bilər?

Tələbə performansına təsir edən amillər kifayət qədər müxtəlifdir və çox vaxt mürəkkəb qarşılıqlı əlaqə təşkil edir. Tələbə performansına təsir edən bəzi əsas amillər bunlardır:

- Motivasiya və maraq- tələbələrin akademik performansını müəyyən edən mühüm amillərdir. Motivasiya tələbənin oxumaq, öyrənmək və məqsədlərinə çatmaq istəyi və əzmidir. Maraq tələbənin müəyyən bir mövzuya və ya mövzuya olan marağı, marağı və həvəsidir. Tələbənin daxili və xarici motivasiyası onun dərslərdə və təlim prosesində iştirakına təsir göstərir. Yüksək öyrənmə motivasiyası tələbəyə təhsil prosesində daha həvəslə iştirak etməyə imkan verir və tələbənin öz qabiliyyətlərinə və uğurlarına inamı onun motivasiyasına təsir edir. Müvəffəqiyyətə güclü inam tələbənin çətinliklərlə üzləşdiyi zaman motivasiyalı olmasına kömək edə bilər. Tələbənin qarşıya qoyduğu məqsədlər və bu məqsədlərə çatmaqda irəliləyişləri onların motivasiyasını artırır. Məqsədlərin aydın, ölçülə bilən və əldə edilə bilən olması vacibdir. Tələbənin dərslərə və ya fənlərə marağı öyrənmə motivasiyasını artırır çünki tələbənin marağına səbəb olan mövzu üzərində işləmək daha çox öyrənmək istəyi yaradır. Tələbənin ətraf mühitlə təsiri də onun motivasiyasına təsir edə bilər. Şagirdin ailəsi, müəllimləri və dostları ilə qarşılıqlı əlaqəsi onların motivasiyasına müsbət və ya mənfi təsir göstərə bilər. Bu amillər şagirdin dərslərdə iştirakını, söylərini və uğurunu şərtləndirən mühüm elementlərdir. Güclü motivasiya və maraq tələbələrə daha effektiv öyrənməyə və uğur qazanmağa imkan verir.
- Tədris keyfiyyəti- tələbələrin təhsil aldıkları mühitin və prosesin keyfiyyətidir. Buraya kurs materialları, tədris metodları və sinif mühiti kimi elementlər daxildir. Yaxşı tədris keyfiyyəti tələbələrin öyrənmə təcrübələrini zənginləşdirir, onların motivasiyasını artırır və uğurlarını təşviq edir. Tədris keyfiyyəti birbaşa olaraq müəllimlərin təcrübə, bilik və tədris bacarıqları ilə bağlıdır. İxtisaslı müəllimlər dərsləri cəlbedici və effektiv şəkildə keçirər, tələbələrə rəhbərlik edə və təlim prosesini təkmilləşdirə bilər. Tədris keyfiyyəti

həmçinin təklif olunan kurs materiallarının keyfiyyəti və müxtəlifliyi ilə ölçülür. Mövcud, uyğun və maraqlı dərsliklər, əyani və audio materiallar şagirdlərin fənləri daha yaxşı başa düşmələrinə və öyrənmələrinə kömək edə bilər. Yaxşı tədris müxtəlif tədris metodları və strategiyalarından istifadə edir. İnteraktiv dərslər, problem həll etmə fəaliyyətləri, qrup işləri, layihələr və sahə tədqiqatları kimi müxtəlif təlim mühitləri tələbələrin öyrənmə təcrübəsini zənginləşdirir. Sınıf mühiti tələbələrin motivasiyasına və öyrənmək istəyinə təsir göstərir. Dəstəkləyici mühit tələbələri özlərini ifadə etməyə, suallar verməyə və yeni ideyalar kəşf etməyə təşviq edir. Yaxşı tədris ədalətli və konstruktiv qiymətləndirmə prosesini təmin edir. Şagirdlərə mütəmadi olaraq rəy verilir və onların nailiyyətləri tanınır. Qiymətləndirmə prosesi tələbələrin inkişafını izləmək və ehtiyaclarını müəyyən etmək üçün istifadə olunur. Texnologiyadan düzgün istifadə tədrisin keyfiyyətini yüksəldə bilər. Təhsil üçün müxtəlif rəqəmsal alətlər və resurslar tələbələrin iştirakını artırır və öyrənməni daha cəlbedici edə bilər.

- Şəxsi amillər- tələbənin fərdi xüsusiyyətləri və şəxsi xüsusiyyətləri ilə əlaqədardır. Bu amillər şagirdin davranışlarına, seçimlərinə və nailiyyətlərinə təsir göstərir. Tələbənin intellekt səviyyəsi və təbii qabiliyyətləri öyrənmə və uğur qazanmasında həlledici rol oynayır. Fərqli tələbələr müxtəlif intellekt və qabiliyyət səviyyələrinə malik ola bilər və bu, onların öyrənmə prosesinə təsir göstərir. Tələbənin öyrənmə üslubu da onların məlumatı anlama və emal etmə üsulunu müəyyən edir. Vizual, eşitmə və ya kinestetik kimi müxtəlif öyrənmə üslubları olan tələbələr fərqli öyrənmə strategiyaları tələb edə bilər. Tələbənin özünə intizam, məsuliyyət hissi, əzmkarlıq və səbr kimi xarakter xüsusiyyətləri şagirdin uğuruna təsir edir. Güclü xarakter tələbənin çətinliklərin öhdəsindən gəlmək və diqqətini məqsədlərinə yönəltmək qabiliyyətini artırır. Tələbənin özünə inamı və uğura inamı öyrənmə prosesinə və fəaliyyətinə təsir göstərir. Güclü özünə inam tələbənin çətinliklərlə üzləşdiyi zaman motivasiyasını saxlamağa kömək edə bilər. Şagirdin öz emosiyalarını, düşüncələrini və davranışlarını idarə etmək bacarığı özünü-tənzimləmə bacarığı

- adlanır. Yaxşı özünütənzimləmə tələbəninin dərslərini və vaxtın idarə edilməsini yaxşılaşdırma bilər.
- Sağlamlıq vəziyyəti- ümumiyyətlə fiziki və psixi sağlamlıqla bağlıdır. Sağlamlıq vəziyyəti tələbəninin diqqətinə, konsentrasiyasına, enerji səviyyəsinə və ümumi rifahına təsir göstərir. Tələbəninin fiziki sağlamlığı onların ümumi sağlamlığına və həyat keyfiyyətinə təsir göstərir. Sağlam bədən daha çox enerji və dözümlü təmin etməklə tələbəyə diqqətini dərslərə yönəltməyə kömək edə bilər. Bundan əlavə, xroniki sağlamlıq problemləri və ya xəstəliklər tələbəninin müntəzəm olaraq dərslərə qatılmasına mane ola bilər və onların fəaliyyətinə mənfi təsir göstərə bilər. Tələbəninin psixi sağlamlığı isə onların idrak funksiyasına və emosional rifahına təsir edir. Psixi sağlamlıq problemləri tələbəninin diqqətini cəmləmə, diqqət və motivasiya qabiliyyətinə mənfi təsir göstərə bilər. Tələbələrin depressiya, narahatlıq və ya stress kimi problemlərin öhdəsindən gəlmək bacarığı onların akademik performansına təsir edə bilər. Tələbələrin stresslə mübarizə qabiliyyəti onların akademik uğurlarına təsir edir. Sıx iş qrafiki, imtahanlar, ev tapşırığı və sosial təzyiqlər tələbələrin stress səviyyələrini artırma bilər. Effektiv stress idarəetmə üsulları tələbəninin stress səviyyəsini azalda və performansını yaxşılaşdırma bilər. Sağlam yuxu və balanslaşdırılmış qidalanma şagirdin diqqətini və konsentrasiyasını yaxşılaşdırma bilər. Qeyri-kafi yuxu və ya qeyri-sağlam qidalanma vərdişləri şagirdin enerji səviyyəsini aşağı sala və öyrənmə prosesinə mənfi təsir göstərə bilər.
 - İqtisadi vəziyyət- tələbəninin təhsil imkanlarına və keyfiyyətinə təsir edə bilər. İqtisadi təhlükəsizlik tələbə uğuruna müsbət təsir göstərə bilər. Bu, tələbəninin iqtisadi vəziyyəti, ailəsinin maddi vəziyyəti və gəlir səviyyəsi ilə bağlıdır. Tələbəninin iqtisadi vəziyyəti ailənin gəlir səviyyəsi, ailənin maddi vəziyyəti və iqtisadi təhlükəsizliyi kimi amillərlə müəyyən edilir. Bu, tələbəninin təhsilə çıxışına, öyrənmə resurslarına çıxışına və ümumi həyat keyfiyyətinə təsir göstərir. Ailənin iqtisadi vəziyyəti tələbəninin təhsil almaq imkanlarına təsir göstərə bilər. Gəlir səviyyəsi yüksək olan ailələrin daha yaxşı təhsil imkanları

olsa da, aztəminatlı ailələrin uşaqları bəzən təhsil imkanlarından məhrum ola bilərlər. İqtisadi vəziyyət tələbələrin təhsil materiallarına və resurslarına çıxışına da təsir göstərir. Yüksək gəlirli ailələr tələbələrini daha çox kitab, kompüter və digər tədris materialları ilə təmin edə bilsələr də, aztəminatlı ailələrin bu cür resurslara çıxışı daha məhdud ola bilər. İqtisadi vəziyyət tələbənin məktəb və kurs seçimini müəyyən edə bilər. Yüksək gəlirli ailələr uşaqlarını özəl məktəblərə göndərə və ya əlavə dərslər keçirə bilsələr də, aztəminatlı ailələrin bu cür seçimlərə çıxışı daha məhdud ola bilər. Aztəminatlı ailələrdə yaşayış şəraiti çox vaxt daha çətin ola bilər ki, bu da tələbənin stress səviyyəsinə təsir edə bilər. İqtisadi təhlükəsizliyin olmaması tələbənin diqqətini yayındıra və təlim prosesinə mənfi təsir göstərə bilər. Aşağı gəlirli ailələrdə tələbələrin işləmək ehtimalı daha yüksək ola bilər. Bu, tələbənin dərslərə və dərslərə vaxt ayırma qabiliyyətinə təsir edə bilər.

1.2 Aşağı performans göstərən tələbələrin üzləşdiyi problemlər

Aşağı performans göstərən tələbələr bir çox problemlərlə üzləşirlər. Bu problemlərə aşağıdakıları misal göstərə bilərik:

- Akademik uğursuzluq: Zəif qiymətlər alan tələbələr kurslarını uğurla başa vurmaqda və buraxılış tələblərini yerinə yetirməkdə çətinliklərlə üzləşə bilərlər. Bu, tələbələrin özünə inamına və motivasiyasına mənfi təsir göstərə bilər.
- Onların akademik karyeralarına təsir: Zəif qiymətlər tələbələrin akademik karyeralarına təsir edə bilər və onların gələcək təhsil və məşğulluq imkanlarını məhdudlaşdıra bilər. Aşağı qiymətlər məzun olduqdan sonra magistr proqramlarına qəbul olmağı çətinləşdirə bilər.
- Təqaüdlərin və dəstəyin itirilməsi: Bəzi tələbələr akademik uğurlarına görə təqaüd və ya maliyyə dəstəyi ala bilərlər. Zəif qiymət alan tələbələr bu cür dəstəyi itirmək riski ilə üzləşə bilər, nəticədə maddi çətinliklər və əlavə stress yarana bilər.

- Psixoloji təsirlər: Zəif qiymətlər tələbələrdə stress, narahatlıq və aşağı özünə hörmət kimi psixoloji problemlər yarada bilər. Bu, tələbələrin emosional və sosial rifahına mənfi təsir göstərə və onların ümumi həyat keyfiyyətini aşağı sala bilər.
- Akademik mühitə təsir: Zəif qiymətlər alan tələbələrin uğursuzluğu qrup dinamikasına və təhsil mühitinə təsir edə bilər. Tələbələrin aşağı motivasiyası və məşğulluğu sinif daxilində qarşılıqlı əlaqəni azalda və tədrisin keyfiyyətini aşağı sala bilər.
- Təhsil müəssisələrinin nüfuzuna xələl gətirə bilər: Zəif qiymət alan tələbələrin yüksək faizinə malik olan təhsil müəssisəsi onun akademik reputasiyasına mənfi təsir göstərə bilər. Bu, universitetin reytingini və qəbul edilən keyfiyyəti aşağı sala bilər və tələbə müraciətinə və qeydiyyat sayına təsir göstərə bilər.

Bu problemlər tələbələrin akademik fəaliyyətini yaxşılaşdırmaq üçün müdaxiləyə ehtiyacı olan sahələri göstərir. Təhsil müəssisələri məsləhət xidmətləri, akademik dəstək proqramları və tələbələrə istiqamət verməklə zəif qiymətləri olan tələbələrə kömək edə və onların uğurlarını artırmağa bilər. Məhz bu səbəblərə görə tələbələrin akademik performansın proqnozlaşdırılması vacib məsələdir. Bu məsələnin həll edilməsində maşın öyrənməsindən istifadə olunur .

1.3 Tələbə müvəffəqiyyətinin proqnozlaşdırılmasının təhsildə önəmi

Tələbə performansını proqnozlaşdırmaq təhsildə bir çox aspektlərdə vacibdir.

Həmin sahələrə aşağıdakıları misal göstərmək olar:

- Tələbə tərəqqisinin monitorinqi: Performans təxminləri tələbələrin akademik tərəqqisini izləmək üçün mühüm vasitədir. Tələbə performansını proqnozlaşdırmaq tələbələrin güclü və zəif tərəflərini müəyyən etməyə kömək edə bilər və müəllimlərə tələbələrə əlavə dəstək göstərməli olduqları yerlər barədə fikir verə bilər.

- Təhsil siyasətlərinin inkişafı: Tələbə performansına dair məlumatlar təhsil siyasətinin inkişafında mühüm rol oynayır. Bu məlumatlar təhsil sistemlərinin güclü və zəif tərəflərini müəyyən etməyə və siyasətçilərə resursları ən səmərəli şəkildə bölüşdürməyə kömək edə bilər.
- Tələbələrə dəstək xidmətlərinə istinad: Performans təxminləri tələbələrə əlavə dəstək xidmətləri göstərməyə kömək edə bilər. Tələbələrin çətinlik çəkdiyi fənlərin müəyyən edilməsi və müvafiq dəstək proqramlarının yaradılması tələbələrin uğurlarını artırır.
- Tədris proseslərinin təkmilləşdirilməsi: Performans proqnozları da tədris proseslərinin təkmilləşdirilməsinə töhfə verə bilər. Tələbələrin fəaliyyətinin qiymətləndirilməsi və tədris metodlarının uyğunlaşdırılması müəllimlərə tələbələri daha effektiv öyrətməyə kömək edə bilər.
- Gələcək uğurun proqnozlaşdırılması: Tələbə performansını proqnozlaşdırmaq tələbələrin gələcək akademik uğurlarını proqnozlaşdırmağa kömək edə bilər. Bu məlumat tələbələrə kollec qəbul perspektivlərini, karyera seçimlərini və karyera yollarını müəyyən etməyə kömək edə bilər.
- Resurslardan səmərəli istifadə: Performans təxminləri təlim resurslarından daha səmərəli istifadə etməyə kömək edə bilər. Bu, tələbələr üçün dəstək proqramlarının və resursların düzgün yerlərə ayrılmasını təmin edə bilər.
- Erkən müdaxilə və istənməyən halların qarşısının alınması: Tələbə performansının proqnozlaşdırılması erkən müdaxilə və qarşısının alınması strategiyalarının hazırlanmasına imkan verir. Tələbələri akademik çətinliklərlə üzləşməzdən əvvəl müəyyən etmək onlara vaxtında dəstək göstərilməsini təmin edə bilər. Bu, tələbə nailiyyətlərini artırmaq və mənfi nəticələrin qarşısını almaq üçün çox vacibdir.
- Fərdiləşdirilmiş öyrənmə yanaşmaları: Tələbə performansını proqnozlaşdırmaq fərdiləşdirilmiş təlim yanaşmalarını inkişaf etdirməyə kömək edə bilər. Tələbələrin güclü və zəif tərəflərinin müəyyən edilməsi müəllimlərə hər bir tələbənin ehtiyaclarına daha yaxşı uyğun gələn tədris strategiyaları hazırlamağa imkan verir.

- Təhsildə bərabərlik və ədalət: Performans təxminləri təhsildə ədalətin və ədalətin təmin edilməsinə töhfə verə bilər. Tələbələrin fəaliyyətinin obyektiv qiymətləndirilməsi tələbələr arasında ədalətliliyi təmin etmək üçün vacibdir. Bu, hər bir tələbəyə öz potensialını reallaşdırmaq üçün bərabər imkanlara çıxışı təmin edə bilər.
- Tələbə motivasiyasının artırılması: Tələbələrin motivasiyasını artırmaq üçün performans proqnozlarından istifadə edilə bilər. Tələbələrin akademik məqsədləri haqqında rəy almaları və gələcək uğurları haqqında öyrənmələri onların təlim prosesinə bağlılığını və motivasiyasını artırır.
- Təhsildə məlumatlara əsaslanan qərarların qəbulu: Performans proqnozları təhsildə məlumatlara əsaslanan qərarların qəbulu proseslərinə kömək edə bilər. Tələbə performansını ilə bağlı məlumatların təhlili məktəb rəhbərlərinə, müəllimlərə və siyasətçilərə məlumat verir və onlara daha məlumatlı qərarlar qəbul etməyə imkan verir.

Bu məqamlar təhsil sistemində müxtəlif aspektlərdə tələbə fəaliyyətinin proqnozlaşdırılmasının vacibliyini daha aydın izah edir. Bu səbəblərə görə tələbə performansının proqnozlaşdırılması təhsil sisteminin effektivliyi və tələbə uğuru üçün kritik əhəmiyyət kəsb edir .

1.4 Maşın öyrənməsinin təhsil sahəsində tətbiqi

Maşın öyrənməsi müxtəlif sənaye sahələrində əhəmiyyətli töhfələr verən təməl qoyma vasitəsi kimi ortaya çıxdı. Təhsil sahəsi də istisna deyil, çünki maşın öyrənmə alqoritmləri tədris və öyrənmə təcrübələrində inqilab etmək üçün istifadə olunur. Məlumatların gücündən və qabaqcıl hesablama texnikalarından istifadə etməklə, maşın öyrənməsi təhsil təcrübələrini dəyişdirmək, təlimatı fərdiləşdirmək və tələbələr üçün təkmilləşdirilmiş nəticələr əldə etmək potensialına malikdir. Təhsildə maşın öyrənməsi müəllimlərə və institutlara tədris yanaşmalarını təkmilləşdirmək üçün innovativ həllər təklif edərək böyük vədlər verir. Geniş məlumat dəstlərini araşdıraraq, maşın öyrənmə alqoritmləri əsas nümunələri müəyyən edə, mənalı

müşahidələr əldə edə və praktiki təkliflər verə bilər. Bu təkliflər daha sonra hər bir tələbənin unikal tələblərinə, güclü və zəif tərəflərinə cavab vermək üçün xüsusi olaraq hazırlanmış fərdiləşdirilmiş təlim trayektoriyalarını yaratmaq üçün istifadə edilə bilər. Nəticə etibarlı ilə, tələbələr dəqiq və diqqətli yardım ala bilər, onlara maksimum imkanlarını açmaq imkanı verir.

Maşın öyrənməsinin təhsildə tətbiq sahələri:

- Maşın öyrənmənin təhsildə fərdiləşdirilmiş öyrənmədə tətbiqi: Maşın öyrənməsi alqoritmləri tələbə performansını, öyrənmə üsulları və üstünlükləri daxil olmaqla böyük həcmdə məlumatları təhlil edə bilər. Bu, hər bir tələbənin ehtiyaclarına uyğunlaşdırılmış fərdiləşdirilmiş təlim yollarının və adaptiv təlimatın yaradılmasına imkan verir. Fərdi güclü və zəif tərəfləri nəzərə alaraq, maşın öyrənmə nişanlanma, motivasiya və ümumi akademik uğuru təşviq edir.
- İntellektual repetitorluq sistemləri: Maşın öyrənməsi real vaxt rejimində geri bildirim və tələbələrə istiqamət verən intellektual repetitorluq sistemlərini gücləndirə bilər. Bu sistemlər tələbə cavablarını və irəliləyişlərini təhlil edir, çətinlik sahələrini müəyyənləşdirir və fərdi müdaxilələr təklif edir. Maşın öyrənmə vasitəsilə repetitorluq sistemləri öz tədris strategiyalarını davamlı olaraq uyğunlaşdırır və təkmilləşdirir, bu da daha effektiv və fərdiləşdirilmiş dəstəyə gətirib çıxarır.
- Avtomatlaşdırılmış qiymətləndirmə və əlaqə: Maşın öyrənmə alqoritmləri qiymətləndirmə proseslərini avtomatlaşdırmağa bilər, müəllimlərin əhəmiyyətli vaxtına və səyinə qənaət edir. Annotasiya edilmiş verilənlər toplusunda modelləri öyrətməklə, maşın öyrənmə sistemləri obyektiv və ardıcıl qiymətləndirmə təmin edərək tapşırıqları, testləri və imtahanları qiymətləndirə bilər. Əlavə olaraq, maşın öyrənmə ilə işləyən rəy sistemləri konstruktiv şərhlər təklif edə bilər, tələbələrə səhvlərini başa düşməyə və performanslarını yaxşılaşdırmağa kömək edir.
- Proqnozlaşdırıcı təhlil: Maşın öyrənmə alqoritmləri məktəbi tərk etmə nisbətləri, akademik uğur və ya məşğulluq səviyyələri kimi tələbə nəticələrini

proqnozlaşdırmaq üçün tarixi məlumatları təhlil edə bilər. Riskli tələbələri erkən müəyyən etməklə, pedaqoqlar məqsədyönlü müdaxilələr və dəstək sistemləri tətbiq edə, məktəbi tərk etmə nisbətlerini azalda və tələbələrin saxlanmasını yaxşılaşdırmağa bilirlər.

Maşın Öyrənməsinin əhəmiyyətini göstərən bəzi əsas məqamlar bunlardır:

- Məlumat istehsalında sürətli artım
- İnsan üçün çətin olan mürəkkəb problemlərin həlli
- Müxtəlif sektorlarda qərarların qəbulu, o cümlədən maliyyə
- Gizli nümunələri tapmaq və məlumatlardan faydalı məlumat çıxarmaq.

1.5 Təhsildə maşın öyrənməsinin tətbiqinin faydaları və çətinlikləri

Təhsildə maşın öyrənməsinin tətbiqinin bir çox faydaları var. Onlardan bəziləri aşağıdakılardır:

- Fərdiləşdirilmiş öyrənmə: Maşın öyrənməsi təlimi fərdi ehtiyac və üstünlüklərə uyğunlaşdırmaqla adaptiv öyrənmə təcrübələrini təmin edir. Bu fərdiləşdirmə tələbələrin iştirakını, motivasiyasını və ümumi öyrənmə nəticələrini artırır.
- Təkmilləşdirilmiş məhsuldarlıq: Maşın öyrənməsi qiymətləndirmə və məlumatların təhlili kimi təkrarlanan tapşırıqları avtomatlaşdırır, müəllimləri daha yüksək səviyyəli təlimatlara diqqət yetirmək üçün azad edir. Bu, təhsil prosesində səmərəliliyi artırır və müəllimlərə tələbələrin dəstəyinə və fərdi təlimatlara daha çox vaxt sərf etməyə imkan verir.
- Məlumata əsaslanan qərarların qəbulu: Maşın öyrənmə alqoritmləri böyük həcmdə məlumatları təhlil edərək, müəllimlər və siyasətçilər üçün dəyərli fikirlər təqdim edir. Nümunələri, tendensiyaları və korrelyasiyaları müəyyən etməklə, maşın öyrənməsi sübuta əsaslanan qərarların qəbulunu dəstəkləyir, bu da qurumlara məqsədyönlü müdaxilələr həyata keçirməyə və daha effektiv təhsil siyasətləri yaratmağa imkan verir.

- Təkmilləşdirilmiş tələbə dəstəyi: Maşın öyrənməsi ilə işləyən sistemlər ani rəy, adaptiv təlimat və fərdiləşdirilmiş müdaxilələr təmin edə bilər. Bu, hər bir tələbənin unikal ehtiyaclarına cavab verən və onların akademik inkişafını dəstəkləyən dəstəkləyici öyrənmə mühitini inkişaf etdirir

Təhsildə maşın öyrənməsinin çətinlikləri:

- Məlumatların keyfiyyəti və məxfilik narahatlıqları: Maşın öyrənmə alqoritmləri dəqiq proqnozlar və fərdiləşdirilmiş tövsiyələr üçün yüksək keyfiyyətli və təmsilçi məlumatlara əsaslanır. Məlumatların məxfiliyinin, təhlükəsizliyinin və etik istifadənin təmin edilməsi etimadı qorumaq və həssas tələbə məlumatlarını qorumaq üçün çox vacibdir.
- Qərəz və ədalət: Maşın öyrənmə alqoritmləri, öyrədildikləri məlumatlarda mövcud olan qərəzləri istəmədən davam etdirə bilər. Bu, mövcud bərabərsizlikləri gücləndirərək ədalətsiz nəticələrə gətirib çıxara bilər. Maşın öyrənmə modellərində qərəzləri aradan qaldırmaq və təhsil təcrübələrində ədaləti və bərabərliyi təmin etmək çox vacibdir.
- Texniki infrastruktur və təhsil: Maşın öyrənmənin təhsildə tətbiqi hardware, proqram təminatı və etibarlı əlaqə daxil olmaqla möhkəm texniki infrastruktur tələb edir. Bundan əlavə, müəllimlərə maşın öyrənmə vasitələrindən səmərəli istifadə etmək üçün adekvat təlimlərin verilməsi sinifə uğurlu inteqrasiya üçün vacibdir.
- İnsan-kompüter qarşılığı: Maşın öyrənməsi dəyərli fikirlər və dəstək təmin edə bilsə də, bu, insan qarşılıqlı əlaqəsini və pedaqoqların təcrübəsini əvəz etməməlidir. Vahid və tələbə mərkəzli öyrənmə təcrübəsini saxlamaq üçün maşın öyrənməsinin istifadəsini insanın iştirakı ilə tarazlaşdırmaq çox vacibdir. Nəticədə, Təhsildə Maşın Öyrənməsi bizim öyrətdiyimiz və öyrəndiyimiz üsulları dəyişdirmək üçün böyük potensiala malikdir. Maşın öyrənmə alqoritmlərindən istifadə etməklə, təhsil müəssisələri öyrənməni fərdiləşdirə, tapşırıqları avtomatlaşdırır, məlumatlara əsaslanan qərarlar qəbul edə və məqsədyönlü tələbə dəstəyi təmin edə bilər. Bununla belə, məlumatların

keyfiyyəti, qərəzlilik, texniki infrastruktur və insan və maşın qarşılıqlı əlaqəsi arasında tarazlığın qorunması kimi problemlər həll edilməlidir. Diqqətli tətbiq və davamlı təkmilləşdirmə yolu ilə Maşın öyrənməsi bütün tələbələr üçün daha effektiv və inklüziv təhsil təcrübələri yaratmağa kömək edə bilər.

II FƏSİL MAŞIN ÖYRƏNMƏSİ NƏDİR?

2.1 Maşın öyrənməsi haqqında məlumat

Real dünyada biz öz təcrübələrimizdən öyrənmə qabiliyyəti ilə hər şeyi öyrənə bilirik və kompüterlər isə bizim təlimatlarımızla müəyyən bir əməliyyat icra edirlər. Sonralar bu maşınların da insan kimi təcrübələrdən və ya keçmiş məlumatlardan öyrənə biləmi kimi bir sual meydana çıxmışdır. Beləliklə, maşın öyrənməsi anlayışı meydana gəlmişdir. Süni intellektin alt dəsti kimi tanınan maşın öyrənməsi ilk növbədə kompüterə verilənlərdən və əvvəlki təcrübələrdən müstəqil şəkildə öyrənməyə imkan verən alqoritmlərin yaradılmasına diqqət yetirir. İlk dəfə olaraq Artur Samuel 1959-cu ildə “maşın öyrənməsi” terminindən istifadə etmişdir. Bu termini aşağıdakı kimi ümumiləşdirmək olar:

Maşın öyrənməsi açıq şəkildə proqramlaşdırılmadan maşına verilənlərdən avtomatik öyrənməyə, təcrübələrdən performansını yaxşılaşdırmağa və hadisələri proqnozlaşdırmağa imkan verir.

Maşın öyrənmə alqoritmləri nümunə tarixi məlumatların və ya təlim məlumatlarının köməyi ilə proqnozlar və ya qərarlar qəbul etməyə kömək edən riyazi model yaradır. Proqnozlaşdırılan modelləri inkişaf etdirmək məqsədilə maşın öyrənməsi kompüter elmini və statistikanı bir araya gətirir. Tarixi məlumatlardan öyrənən alqoritmlər ya qurulur, ya da maşın öyrənməsində istifadə olunur. Alqoritmin performansını təmin etdiyimiz məlumatların miqdarına mütənasib olaraq yüksələcək. Maşın, performansını yaxşılaşdırmaq üçün daha çox əldə edə biləcəyi məlumatı öyrənə bilər. Maşın öyrənmə sistemi proqnozlaşdırma modelləri qurur, əvvəlki məlumatlardan öyrənir və hər dəfə yeni məlumatların çıxışını proqnozlaşdırır. Verilənlərin ölçüsü çıxışı dəqiq proqnozlaşdırıran daha yaxşı model yaratmağa kömək edir, bu da öz növbəsində proqnozlaşdırılan çıxışın düzgünlüyünə təsir edir.

Maşın öyrənməsinin zəruriliyinə nümunə kimi mürəkkəb bir problemlə bağlı proqnoz verməli olduğunu göstərək. Bu cür problemlərin həllində kod yazmaq əvəzinə, biz sadəcə olaraq məlumatları ümumi alqoritmlərə öyrətməliyik, hansı ki,

həmin alqoritmlər verilənlər əsasında məntiqi əlaqə qurur və nəticəni proqnozlaşdırır.

Maşın öyrənməsi xüsusiyyətləri bunlardır:

- Maşın öyrənməsi verilənlər toplusunda müxtəlif nümunələri aşkar etmək üçün verilənlərdən istifadə edir.
- O, keçmiş məlumatlardan öyrənə və avtomatik təkmilləşə bilər.
- Maşın öyrənməsi verilənlərə əsaslanan texnologiyadır.
- O, data minigə yəni, verilənlərin əldə edilməsinə çox bənzəyir, çünki o, həm də böyük ölçüdə məlumatla işləyə bilir.

2.2 Maşın öyrənməsinin tətbiq sahələri

Maşın öyrənməsinə tələbat son zamanlarda daha çox artır. Bir insanın birbaşa həyata keçirə bilməyəcəyi çox mürəkkəb vəzifələri yerinə yetirə bildiyi üçün maşın öyrənməsi bir çox sahələrdə tələb olunur. İnsanlar böyük həcmdə məlumatların idarə edilməsində çətinliklərlə üzləşdiyinə görə maşın öyrənməsindən istifadə edilən kompüter sistemlərinə ehtiyac duyulur. Maşın öyrənməsi gündən-günə çox sürətlə inkişaf edir. Google Maps, Google assistant, Alexa və s. kimi vasitələrlə gündəlik həyatımızda maşın öyrənməsindən istifadə edirik. Aşağıda tətbiq olunduğu sahələrdən bəziləri qeyd olunmuşdur:

- Təsvirin tanınması: Maşın öyrənməsinin ən geniş yayılmış tətbiqlərindən biri təsvirin tanınmasıdır. O, əsasən obyektləri, şəxsləri, yerləri, rəqəmsal şəkilləri və s. müəyyən etmək üçün istifadə olunur. Şəkilin tanınması və üz aşkarlanmasının məşhur istifadə halına Avtomatik dost işarələmə təklifini misal göstərə bilərik. Beləki, Facebook bizə avtomatik dost işarələmə təklifi funksiyasını təqdim edir. Facebook dostlarımızla şəkil yüklədiyimiz zaman avtomatik olaraq adla etiketləmə təklifi edir və bunun arxasında duran texnologiya məhz maşın öyrənməsinin üz tanıma alqoritmidir. O, şəkildəki sifətin tanınması və şəxsiyyətin identifikasiyasına cavabdeh olan "Deep Face" adlı Facebook layihəsindən istifadə edir.

- Nitqin tanınması: Google-dan istifadə edərkən biz "Səslə axtarış" seçimi əldə edirik, o, nitqin tanınması vasitəsilə olur və bu, maşın öyrənməsinin məşhur tətbiqlərindən biridir. Nitqin tanınması səsli göstərişlərin mətnə çevrilməsi prosesidir və bu proses, "Nitqdən mətnə" və ya "Kompüter nitqinin tanınması" kimi də tanınır. Hazırda maşın öyrənməsi alqoritmləri nitqin tanınmasının müxtəlif tətbiqləri tərəfindən geniş istifadə olunur. Məsələn, Google köməkçisi, Siri, Cortana və Alexa səsli təlimatlara əməl etmək üçün nitqin tanınması texnologiyasından istifadə edən texnologiyalardır.
- Trafik proqnozu: Yeni bir yerə baş çəkmək istədikdə, bizə ən qısa marşrutla düzgün yolu göstərən və nəqliyyat şəraitini proqnozlaşdıran Google Xəritələrin köməyindən istifadə edirik. Bu texnologiya, iki yolun köməyi ilə nəqliyyatın boşaldılması, yavaş hərəkət etməsi və ya sıx tıxac olması kimi yol şəraitini proqnozlaşdırır. Google Xəritədən istifadə edən hər kəs bu tətbiqin daha yaxşı olmasına kömək edir. Beləki, sistem mütamadi olaraq istifadəçidən məlumat alır və performansını yaxşılaşdırmaq üçün verilənlər bazasına göndərir.
- Məhsul tövsiyələri: Maşın öyrənməsi istifadəçiyə məhsul tövsiyəsi üçün bir çox müxtəlif e-ticarət və əyləncə şirkətləri tərəfindən geniş istifadə olunur. Məsələn, Amazonda hansısa məhsulu axtardığımız zaman eyni brauzerdə internetdə gəzərkən eyni məhsulun reklamını almağa başlayırıq və bunun səbəbi isə maşın öyrənməsidir.
- Özünü idarə edən avtomobillər: Maşın öyrənmənin maraqlı tətbiqlərindən biri özünü idarə edən avtomobillərdir. Maşın öyrənməsi özü idarə olunan avtomobillərdə mühüm rol oynayır. Ən məşhur avtomobil istehsalçısı olan Tesla sürücüsüz avtomobil üzərində işləyir. Sürücülük zamanı insanları və obyektləri aşkar etmək üçün avtomobil modellərini öyrətmək üçün isə sistem nəzarətsiz öyrənmə metodundan istifadə edir.
- E-poçt spam və zərərli proqram filtrləmə: Hər dəfə yeni e-poçt aldıqda, o, avtomatik olaraq vacib, normal və spam olaraq süzülür. Biz həmişə spam qutumuzda vacib simvol və spam e-poçtları olan vacib məktubları gələnlər

qutumuzda alırıq və bunun arxasındakı texnologiya Maşın öyrənməsidir. Məsələn, Gmail tərəfindən məzmun filtri, başlıq filtri, ümumi qara siyahı filtri, qaydalara əsaslanan filtrlər, icazə filtrləri və s. kimi spam filterlərindən istifadə edilir. Bu zaman Layer Perceptron, Qərar ağacı və Naive Bayes təsnifatı kimi bəzi maşın öyrənmə alqoritmləri e-poçt spamlarının filtrasiyası və zərərli proqramların aşkarlanması üçün istifadə olunur.

- Virtual şəxsi köməkçi: Artıq Google köməkçisi, Alexa, Cortana, Siri kimi müxtəlif virtual şəxsi köməkçilərimiz var. Adından göründüyü kimi, onlar səsli təlimatımızdan istifadə edərək məlumatı tapmaqda bizə kömək edirlər. Bu köməkçilər musiqi çalmaq, kiməsə zəng etmək, e-poçt açmaq, görüş təyin etmək və s. kimi səsli göstərişlərimizlə bizə müxtəlif yollarla kömək edirlər. Bu köməkçi səs təlimatlarımızı qeyd edir, onu buludda server üzərindən göndərir və maşın öyrənmə alqoritmlərindən istifadə edərək kodunu açır və buna uyğun hərəkət edir.
- Onlayn fırıldaqçılığın aşkarlanması: Maşın öyrənməsi fırıldaqçılıq əməliyyatını aşkar edərək onlayn əməliyyatımızı daha təhlükəsiz edir. Hər hansı bir onlayn əməliyyat həyata keçirdikdə, saxta hesablar, saxta idlər və əməliyyatın ortasında pul oğurlamaq kimi saxta əməliyyatın baş verə biləcəyi müxtəlif hallar ola bilər. Bu zaman bunu aşkar etmək üçün Feed Forward Neyron şəbəkəsi bunun həqiqi bir əməliyyat və ya fırıldaq əməliyyatı olub olmadığını yoxlamaqla bizə kömək edir. Hər bir orijinal əməliyyat üçün çıxış bəzi hash dəyərlərinə çevrilir və bu dəyərlər növbəti dövr üçün girişə çevrilir. Hər bir orijinal əməliyyat üçün və fırıldaqçılıq əməliyyatı üçün dəyişiklik əldə edən xüsusi bir model var. Bu model onu aşkar edir və onlayn əməliyyatlarımızı daha təhlükəsiz edir.
- Birja ticarəti: Maşın öyrənməsi birja ticarətində geniş istifadə olunur. Birjada səhmlərdə hər zaman yüksəliş və eniş riski var, ona görə də bu maşın öyrənməsi üçün uzunmüddətli və qısamüddətli yaddaş neyron şəbəkəsi fond bazarı tendensiyalarının proqnozlaşdırılması üçün istifadə olunur.

- Tibbi diaqnoz: Tibb elmində maşın öyrənməsi xəstəliklərin diaqnozu üçün geniş şəkildə istifadə olunur. Bununla, tibbi texnologiya çox sürətlə inkişaf edir və hətta beyindəki lezyonların dəqiq mövqeyini proqnozlaşdırma bilən 3D modellər qura bilir və bununlada beyin şişlərini və beyinlə əlaqəli digər xəstəlikləri asanlıqla tapmağa kömək edir.
- Avtomatik dil tərcüməsi: İstənilən bir dildə tərcümə etmək zərurəti yarandıqda maşın öyrənməsi mətni məlum dillərimizə çevirməklə bizə kömək edir. Bu prosesi həyata keçirən Google-un mətni bizim tanış dilimizə tərcümə edən Neyron Maşın Öyrənməsidir və buna avtomatik tərcümə deyilir. Avtomatik tərcümənin arxasında duran texnologiya və təsvirin tanınması üçün istifadə olunan və mətni bir dildən digər dilə tərcümə edən ardıcılıqla öyrənmə alqoritmidir.
- Kibertəhlükəsizlik: Maşın öyrənməsi eyni zamanda kiberhücumları aşkar etmək və qarşısını almaq üçün sistemlər yaratmaq üçün istifadə olunur. Bunlar bizim həyatımızda və texnologiyanın gələcəyində əhəmiyyətini nümayiş etdirəcək maşın öyrənməsinin tətbiqlərindən bir neçə nümunədir. Maşın öyrənməsindən istifadə etməklə həm vaxta, həm də pula qənaət edə bilərik.

2.3 Maşın öyrənməsinin təsnifatı

Geniş səviyyədə maşın öyrənməsi üç növə təsnif edilə bilər:

- Nəzarət olunan öyrənmə
- Nəzarətsiz öyrənmə
- Möhkəmləndirici öyrənmə

Nəzarət olunan öyrənmə, maşınların yaxşı "etiketli" təlim məlumatlarından istifadə edərək öyrədildiyi maşın öyrənməsi növüdür və bu məlumatlar əsasında maşınlar nəticəni proqnozlaşdırır. Etiketlənmiş məlumatlar bəzi giriş məlumatlarının artıq düzgün çıxışla işarələndiyini bildirir. Nəzarət olunan öyrənmədə maşınlara verilən təlim məlumatları maşınlara çıxışı düzgün proqnozlaşdırmağı öyrədən nəzarətçi kimi işləyir. O, tələbənin müəllimin nəzarətində öyrəndiyi anlayışı tətbiq

edir. Nəzarət edilən öyrənmə, maşın öyrənmə modelinə daxilolma məlumatı və düzgün çıxış məlumatının təmin edilməsi prosesidir. Nəzarət olunan öyrənmə alqoritminin məqsədi giriş dəyişənini (x) çıxış dəyişəni (y) ilə xəritələşdirmək üçün xəritəçəkmə funksiyasını tapmaqdır. Real dünyada nəzarət edilən öyrənmə riskin qiymətləndirilməsi, şəkillərin təsnifatı, fırıldaqçılığın aşkarlanması, spam filtrasiyası və s. üçün istifadə edilə bilər.

Nəzarət olunan öyrənmədə modellər etiketli verilənlər toplusundan istifadə etməklə öyrədilir, burada model hər bir məlumat növü haqqında öyrənir. Təlim prosesi başa çatdıqdan sonra model test məlumatları əsasında sınaqdan keçirilir (təlim toplusunun alt dəsti) və sonra o, nəticəni proqnozlaşdırır. Nəzarət altında olan təlimi daha iki növ problemə bölmək olar:

1. Reqressiya: Giriş dəyişəni ilə çıxış dəyişəni arasında əlaqə olduqda reqressiya alqoritmlərindən istifadə edilir. O, hava proqnozu, bazar trendləri və s. kimi davamlı dəyişənlərin proqnozlaşdırılması üçün istifadə olunur. Reqressiya alqoritmlərinə xətti reqressiya, reqressiya ağacları, qeyri-xətti reqressiya, Bayes xətti reqressiyası, polinom reqressiyası kimi məşhur alqoritmləri misal göstərmək olar.
2. Təsnifat: Təsnifat alqoritmı təlim məlumatları əsasında yeni müşahidələrin kateqoriyasını müəyyən etmək üçün istifadə edilən Nəzarət Edilən Öyrənmə texnikasıdır. Təsnifatda proqram verilmiş verilənlər toplusundan və ya müşahidələrdən öyrənir və sonra yeni müşahidəni bir sıra siniflərə və ya qruplara təsnif edir. Reqressiyadan fərqli olaraq, Təsnifatın çıxış dəyişəni "doğru və ya yanlış" kimi dəyər deyil, kateqoriyadır. Təsnifat alqoritmı nəzarət olunan öyrənmə texnikası olduğundan, etikətlənmiş giriş məlumatlarını qəbul edir. Bu o deməkdir ki, o, müvafiq çıxışı olan girişi ehtiva edir. Təsnifat alqoritmində diskret çıxış funksiyası- y , giriş dəyişəninə - x uyğunlaşdırılır.

Reqressiya və təsnifat alqoritmləri arasındakı fərqlər müxtəlif fərqlər var:

- Reqressiyada çıxış dəyişəni davamlı xarakterli və ya real dəyərdə olmalıdır. Təsnifatda isə çıxış dəyişəni diskret dəyər olmalıdır.
- Reqressiya alqoritminin vəzifəsi giriş dəyərini (x) davamlı çıxış dəyişəni (y) ilə xəritələşdirməkdir. Təsnifat alqoritminin vəzifəsi giriş dəyərini(x) diskret çıxış dəyişəni (y) ilə xəritələşdirməkdir.
- Reqressiya alqoritmləri həqiqi verilənlərdən istifadə olunur. Təsnifat alqoritmləri diskret verilənlərlə istifadə olunur.
- Reqressiyada biz çıxışı daha dəqiq proqnozlaşdırmağa bilən ən yaxşı uyğunluq xəttini tapmağa çalışırıq. Təsnifatda isə verilənlər toplusunu müxtəlif siniflərə bölə bilən qərar sərhədini tapmağa çalışırıq.
- Reqressiya alqoritmini daha da xətti və qeyri-xətti reqressiyaya bölmək olar. Təsnifat alqoritmlərini İkili təsnifat və çox sinif təsnifatına bölmək olar. Dissertasiyada əsas məqsəd ümumi orta müvəffəqiyyət göstəricilərini proqnozlaşdırmaq olduğu üçün nəzarət olunan öyrənmə alqoritmlərinin reqressiyaya növündən istifadə edilir.

Nəzarətli öyrənmənin üstünlükləri:

- Nəzarət olunan öyrənmənin köməyi ilə model əvvəlki təcrübələr əsasında nəticəni proqnozlaşdırmağa bilər.
- Nəzarətli təlimdə biz obyektlərin sinifləri haqqında dəqiq təsəvvürə malik ola bilərik.
- Nəzarət edilən öyrənmə modeli bizə saxtakarlığın aşkarlanması, spam filtrasiyası və s. kimi müxtəlif real problemləri həll etməyə kömək edir.

Nəzarətli öyrənmənin çatışmazlıqları:

- Nəzarət olunan öyrənmə modelləri mürəkkəb tapşırıqların həlli üçün uyğun deyil.
- Nəzarət edilən öyrənmə test məlumatları təlim məlumat dəstindən fərqli olarsa, düzgün nəticəni proqnozlaşdırmağa bilməz.
- Təlim çoxlu hesablama vaxtı tələb edirdi.

- Nəzarətli təlimdə obyektin sinifləri haqqında kifayət qədər biliyə ehtiyacımız var.

Nəzarətsiz öyrənmə: Bir çox hallar ola bilər ki, bizdə etiketlenmiş məlumat olmaya bilər və verilmiş verilənlər toplusundan gizli nümunələri tapmaq lazımdır. Beləliklə, maşın öyrənməsində bu cür halları həll etmək üçün bizə nəzarətsiz öyrənmə üsulları lazımdır. Nəzarətsiz öyrənmə nədir? Adından da görüldüyü kimi, nəzarətsiz öyrənmə, təlim verilənlər bazasından istifadə edərək modellərə nəzarət olunmayan maşın öyrənmə texnikasıdır. Bunun əvəzinə, modellər verilmiş məlumatlardan gizli nümunələri və anlayışları tapır. Bunu yeni şeylər öyrənərkən insan beynində baş verən öyrənmə ilə müqayisə etmək olar. Bu kimi müəyyən edilə bilər: Nəzarətsiz öyrənmə, modellərin etiketlenməmiş verilənlər toplusundan istifadə etməklə öyrədildiyi və heç bir nəzarət olmadan həmin data üzərində işləməsinə icazə verilən maşın öyrənməsi növüdür. Nəzarətsiz öyrənmə birbaşa reqressiya və ya təsnifat probleminə tətbiq edilə bilməz, çünki nəzarət edilən öyrənmədən fərqli olaraq bizdə giriş məlumatı var, lakin müvafiq çıxış məlumatları yoxdur. Nəzarətsiz öyrənmənin məqsədi verilənlər toplusunun əsas strukturunu tapmaq, həmin məlumatları oxşarlıqlara görə qruplaşdırmaq və həmin verilənlər toplusunu sıxılmış formatda təqdim etməkdir. Alqoritm heç vaxt verilmiş verilənlər toplusu üzərində öyrədilmir, yəni verilənlər bazasının xüsusiyyətləri haqqında heç bir fikri yoxdur. Nəzarətsiz öyrənmə alqoritminin vəzifəsi təsvirin xüsusiyyətlərini özbaşına müəyyən etməkdir. Nəzarətsiz öyrənmə alqoritmi bu tapşırıqı şəkillər arasındakı oxşarlıqlara uyğun olaraq təsvir verilənlər toplusunu qruplara salmaqla yerinə yetirəcək.

Aşağıda nəzarətsiz öyrənmənin əhəmiyyətini təsvir edən bəzi əsas səbəblər verilmişdir:

- Nəzarətsiz öyrənmə məlumatlardan faydalı fikirlər tapmaq üçün faydalıdır.
- Nəzarətsiz öyrənmə, insanın öz təcrübələri ilə düşünməyi öyrənməsi ilə çox oxşardır, bu da onu həqiqi süni intellektə yaxınlaşdırır.
- Nəzarətsiz öyrənməni daha vacib edən etiketsiz və kateqoriyalaşdırılmamış məlumatlar üzərində işləyir.

- Real həyatda biz həmişə müvafiq çıxışı olan giriş məlumatlarına malik deyilik, ona görə də belə halları həll etmək üçün bizə nəzarətsiz öyrənmə lazımdır.

Burada etiketsiz bir giriş məlumatı götürdük, yəni o kateqoriyalara bölünməyib və müvafiq çıxışlar da verilməyib. İndi bu etiketsiz giriş məlumatları onu öyrətmək üçün maşın öyrənmə modelinə verilir. Birincisi, o, məlumatlardan gizli nümunələri tapmaq üçün xam məlumatları şərh edəcək və sonra k-vasitələri klasterləşdirmə, Qərar ağacı və s. kimi uyğun alqoritmləri tətbiq edəcəkdir. Uyğun alqoritmi tətbiq etdikdən sonra alqoritm verilənlər obyektlərini obyektlər arasındakı oxşarlıq və fərqlərə görə qruplara bölür. Nəzarətsiz öyrənmə alqoritminin növləri: klasterləşdirmə və assosiasiya

Klasterləşdirmə: Klasterləşdirmə obyektlərin çoxluqlara qruplaşdırılması üsuludur ki, ən çox oxşarlığı olan obyektlər qrup halında qalsın və digər qrupun obyektləri ilə daha az və ya heç oxşarlığı yoxdur. Klaster təhlili məlumat obyektləri arasında ümumi cəhətləri tapır və onları bu ümumi cəhətlərin mövcudluğu və olmamasına görə təsnif edir.

Assosiasiya: Assosiasiya qaydası, böyük verilənlər bazasında dəyişənlər arasında əlaqələri tapmaq üçün istifadə edilən nəzarətsiz öyrənmə metodudur. Verilənlər dəstində birlikdə baş verən elementlər dəstini müəyyən edir. Assosiasiya qaydası marketinq strategiyasını daha effektiv edir. Məsələn, X məhsulu alan insanlar da Y məhsulunu almağa meyillidirlər. Assosiasiya qaydasının tipik nümunəsi bazar səbətinin təhlilidir.

Nəzarətsiz öyrənmə alqoritmlərinə K-vasitəsi klasterləşdirmə, KNN (k-ən yaxın qonşular), İyerarxal klasterləşmə, Anomaliya aşkarlanması, Neyron şəbəkələri, Prinsip Komponent Təhlili, Müstəqil Komponent Təhlili, Apriori alqoritmi, Tək dəyər parçalanması alqoritmləri aiddir.

Nəzarətsiz öyrənmənin üstünlükləri:

- Nəzarətsiz öyrənmə nəzarət edilən öyrənmə ilə müqayisədə daha mürəkkəb tapşırıqlar üçün istifadə olunur, çünki nəzarətsiz öyrənmədə etiketli daxiletmə datamız yoxdur.
- Nəzarətsiz öyrənməyə üstünlük verilir, çünki etiketli məlumatlarla müqayisədə etiketsiz məlumat əldə etmək asandır.

Nəzarətsiz öyrənmənin çatışmazlıqları:

- Nəzarətsiz öyrənmə nəzarət edilən öyrənmə ilə müqayisədə mahiyyətə daha çətindir, çünki onun müvafiq nəticəsi yoxdur.
- Nəzarətsiz öyrənmə alqoritminin nəticəsi daha az dəqiq ola bilər, çünki giriş məlumatları etikətlənməmişdir və alqoritmlər əvvəlcədən dəqiq çıxışı bilmirlər.

Nəzarət olunan və nəzarətsiz öyrənmə maşın öyrənməsinin iki üsuludur. Ancaq hər iki üsul fərqli ssenarilərdə və fərqli məlumat dəstləri ilə istifadə olunur. Nəzarət olunan və nəzarətsiz öyrənmə alqoritmləri arasında fərqlər bulardır:

- Nəzarət olunan öyrənmə alqoritmləri etikətlənmiş məlumatlardan istifadə etməklə öyrədilir. Lakin nəzarətsiz öyrənmə alqoritmləri etiketsiz verilənlərdən istifadə etməklə öyrədilir.
- Nəzarət olunan öyrənmə modeli düzgün çıxışı proqnozlaşdırıb-proqnozlaşdırmadığını yoxlamaq üçün birbaşa rəy alır. Nəzarətsiz öyrənmə modeli isə heç bir rəy qəbul etmir.
- Nəzarət olunan öyrənmə modeli çıxışı proqnozlaşdırır. Nəzarətsiz öyrənmə modeli isə verilənlərdə gizli nümunələri tapır.
- Nəzarət olunan öyrənmədə modelə giriş məlumatları çıxışla birlikdə verilir. Ancaq Nəzarətsiz öyrənmədə modelə yalnız giriş məlumatları verilir.
- Nəzarət olunan öyrənmənin məqsədi modeli öyrətməkdir ki, ona yeni məlumatlar verildikdə nəticəni proqnozlaşdırma bilsin. Nəzarətsiz öyrənmənin məqsədi isə naməlum verilənlər bazasından gizli nümunələri və faydalı fikirləri tapmaqdır.

- Nəzarət olunan öyrənmə modeli öyrətmək üçün nəzarətə ehtiyac duyar. Nəzarətsiz öyrənmə modeli öyrətmək üçün heç bir nəzarətə ehtiyac duymur.
- Nəzarət olunan öyrənmə təsnifat və reqressiya problemlərində təsnif edilə bilər. Nəzarətsiz öyrənmə klasterləşdirmə və assosiasiya problemlərində təsnif edilə bilər.
- Nəzarət olunan öyrənmə, giriş və müvafiq çıxışları bildiyimiz hallarda istifadə edilə bilər. Nəzarətsiz öyrənmə bizdə yalnız giriş məlumatlarının olduğu və müvafiq çıxış məlumatlarının olmadığı hallar üçün istifadə edilə bilər.
- Nəzarət edilən öyrənmə modeli dəqiq nəticə verir. Nəzarətsiz öyrənmə modeli nəzarət edilən öyrənmə ilə müqayisədə daha az dəqiq nəticə verə bilər.
- Nəzarət edilən öyrənmə əsl Süni intellektə yaxın deyil, çünki burada biz əvvəlcə hər bir məlumat üçün modeli öyrədirik, sonra isə yalnız düzgün çıxışı proqnozlaşdırma bilər. Nəzarətsiz öyrənmə əsl Süni İntellektə daha yaxındır, çünki öz təcrübələri əsasında yeni məlumatlar öyrənir.

Gücləndirici öyrənmə: Gücləndirici öyrənmə agentin hərəkətləri yerinə yetirərək və hərəkətlərin nəticələrini görərək mühitdə davranmağı öyrəndiyi əks əlaqəyə əsaslanan maşın öyrənmə texnikasıdır. Hər yaxşı hərəkət üçün agent müsbət rəy alır və hər bir pis hərəkət üçün agent mənfi rəy və ya cəza alır. Gücləndirici öyrənmədə agent nəzarət edilən öyrənmədən fərqli olaraq heç bir etiketli məlumat olmadan əks əlaqədən istifadə edərək avtomatik öyrənir. Heç bir etiketli məlumat olmadığı üçün agent yalnız öz təcrübəsi ilə öyrənməyə borcludur. Gücləndirici öyrənmə qərar vermənin ardıcıl olduğu və məqsədin uzunmüddətli olduğu müəyyən bir problem növünü həll edir, məsələn, oyun oynamaq, robototexnika və s. Agent ətraf mühitlə qarşılıqlı əlaqədə olur və onu özü araşdırır. Gücləndirici öyrənmədə agentin əsas məqsədi maksimum müsbət mükafatlar əldə etməklə performansını artırmaqdır. Agent zərbə və sınaq prosesi ilə öyrənir və təcrübəyə əsaslanaraq tapşırığı daha yaxşı şəkildə yerinə yetirməyi öyrənir. Beləliklə, deyə bilərik ki, "Gücləndirici öyrənmə, ağıllı agentin (kompüter proqramı) ətraf mühitlə qarşılıqlı əlaqədə olduğu və bunun çərçivəsində hərəkət etməyi öyrəndiyi bir maşın öyrənmə metodu növüdür." Robot

itin qollarının hərəkətini necə öyrənməsi gücləndirici öyrənmə nümunəsidir. Gücləndirici öyrənmə süni intellektin əsas hissəsidir və bütün süni intellekt agentləri gücləndirici öyrənmə konsepsiyası üzərində işləyir. Burada agenti əvvəlcədən proqramlaşdırmağa ehtiyac yoxdur, çünki o, heç bir insan müdaxiləsi olmadan öz təcrübəsindən öyrənir. Misal: Tutaq ki, labirint mühitində süni intellekt agentləri var və onun məqsədi müəyyən bir obyekt tapmaqdır. Agent bəzi hərəkətləri yerinə yetirərək ətraf mühitlə qarşılıqlı əlaqə qurur və bu hərəkətlərə əsasən agentin vəziyyəti dəyişir, həmçinin rəy olaraq mükafat və ya cəza alır. Agent bu üç şeyi etməyə davam edir (hərəkət etmək, vəziyyəti dəyişmək və ya eyni vəziyyətdə qalmaq və rəy almaq) və bu hərəkətləri etməklə o, ətraf mühiti öyrənir və araşdırır. Agent öyrənir ki, hansı hərəkətlər müsbət rəy və ya mükafatlara səbəb olur və hansı hərəkətlər mənfi rəy cəzasına səbəb olur. Müsbət mükafat olaraq agent müsbət xal alır, cəza olaraq isə mənfi xal alır.

Gücləndirici öyrənmədə istifadə olunan terminlər bunlardır:

- Agent: Ətraf mühiti dərk edən/tədqiq edən və ona uyğun hərəkət edə bilən komputer proqramıdır.
- Mühit : Agentin mövcud olduğu və ya əhatə olunduğu vəziyyət. RL-də biz stoxastik mühiti qəbul edirik, yəni təsadüfi təbiətdir.
- Fəaliyyət : Fəaliyyətlər agent tərəfindən mühitdə edilən hərəkətlərdir.
- State: State agent tərəfindən görülən hər bir hərəkətdən sonra mühit tərəfindən qaytarılan vəziyyətdir.
- Mükafat: Agentin fəaliyyətini qiymətləndirmək üçün mühitdən agentə qaytarılan rəy.
- Siyasət: Siyasət agent tərəfindən cari vəziyyətə əsaslanan növbəti fəaliyyət üçün tətbiq edilən strategiyadır.
- Dəyər : Endirim faktoru ilə uzunmüddətli və qısamüddətli mükafatın əksinə olaraq yenidən tənzimlənməsi gözlənilir.
- Q-value: O, əsasən dəyərə bənzəyir, lakin cari hərəkət kimi əlavə bir parametrlər alır.

Gücləndirici öyrənmənin əsas xüsusiyyətləri bunlardır:

- Gücləndirici öyrənmədə agentə ətraf mühit və hansı tədbirlərin görülməsi lazım olduğu barədə təlimat verilmir.
- Agent növbəti hərəkəti edir və əvvəlki hərəkətin rəyinə uyğun olaraq vəziyyəti dəyişir.
- Agent gecikmiş mükafat ala bilər.
- Mühit stoxastikdir və agent maksimum müsbət mükafatlar əldə etmək üçün onu araşdırmalıdır.

Gücləndirici öyrənmənin həyata keçirilməsi üçün yanaşmalar müxtəlifdir. Maşın öyrənməsində gücləndirici öyrənmənin həyata keçirilməsinin əsasən üç yolu var, bunlar aşağıdakılardır:

1) Dəyər əsaslı: Dəyər əsaslı yanaşma hər hansı bir siyasət altında bir vəziyyətdə maksimum dəyər olan optimal dəyər funksiyasını tapmağa əsaslanır.

2) Siyasət əsaslı: Siyasət əsaslı yanaşma dəyər funksiyasından istifadə etmədən gələcək maksimum mükafatlar üçün optimal siyasəti tapmaqdır. Bu yanaşmada agent elə bir siyasət tətbiq etməyə çalışır ki, hər addımda yerinə yetirilən hərəkət gələcək mükafatı maksimum dərəcədə artırmağa kömək etsin. Siyasət əsaslı yanaşma əsasən iki növ siyasətə malikdir:

- Deterministik: Eyni hərəkət istənilən vəziyyətdə siyasət (π) tərəfindən istehsal olunur.
- Stokastik: Bu siyasətdə ehtimal istehsal olunan hərəkəti müəyyən edir.

2) Model əsaslı: Modelə əsaslanan yanaşmada mühit üçün virtual model yaradılır və agent onu öyrənmək üçün həmin mühiti araşdırır. Bu yanaşma üçün heç bir xüsusi həll və ya alqoritm yoxdur, çünki model təqdimatı hər bir mühit üçün fərqlidir.

2.4 Maşın öyrənmə alqoritmləri

Regressiya təhlili bir və ya bir neçə müstəqil dəyişənə malik asılı (hədəf) və müstəqil (proqnozlaşdırıcı) dəyişənlər arasındakı əlaqəni modelləşdirmək üçün statistik üsuldur. Daha dəqiq desək, regressiya təhlili, digər müstəqil dəyişənlər sabit saxlanıldıqda asılı dəyişənin dəyərinin müstəqil dəyişənə uyğun olaraq necə dəyişdiyini anlamağa kömək edir. Temperatur, yaş, əmək haqqı, qiymət və s. kimi davamlı, real dəyərləri proqnozlaşdırır. Regressiya, dəyişənlər arasında korrelyasiya tapmağa kömək edən və bir və ya bir neçə proqnozlaşdırıcı dəyişənə əsaslanaraq davamlı çıxış dəyişənini proqnozlaşdırmağa imkan verən nəzarət edilən öyrənmə texnikasıdır. Əsasən proqnozlaşdırma, proqnozlaşdırma, zaman sıralarının modelləşdirilməsi və dəyişənlər arasında səbəb-nəticə əlaqəsinin müəyyən edilməsi üçün istifadə olunur.

Regressiyada biz verilən məlumat nöqtələrinə ən yaxşı uyğun gələn dəyişənlər arasında qrafik çəkirik, bu sxemdən istifadə edərək maşın öyrənmə modeli verilənlər haqqında proqnozlar verə bilər. Sadə sözlərlə desək, regressiya, məlumat nöqtələri ilə regressiya xətti arasındakı şaquli məsafənin minimum olması üçün hədəf-proqnozlaşdırıcı qrafikdəki bütün məlumat nöqtələrindən keçən xətt və ya əyri göstərir. Məlumat nöqtələri və xətt arasındakı məsafə modelin güclü əlaqə qurduğunu və ya olmadığını bildirir. Regressiyanın bəzi nümunələri aşağıdakı kimi ola bilər:

- Temperaturdan və digər amillərdən istifadə edərək yağışın proqnozlaşdırılması
- Bazar meyllərinin müəyyən edilməsi
- Tələsik sürücülük səbəbindən yol qəzalarının proqnozu.

Regressiya təhlili ilə bağlı terminologiyalar bunlardır:

- Asılı dəyişən: Regressiya təhlilində proqnozlaşdırmaq və ya anlamaq istədiyimiz əsas amil asılı dəyişən adlanır. Buna hədəf dəyişən də deyilir.
- Müstəqil Dəyişən: Asılı dəyişənlərə təsir edən və ya asılı dəyişənlərin qiymətlərini proqnozlaşdırmaq üçün istifadə olunan amillərə müstəqil dəyişən, həmçinin proqnozlaşdırıcı da deyilir.

- Outliers: Outlier, digər müşahidə edilən dəyərlərlə müqayisədə ya çox aşağı dəyər, ya da çox yüksək dəyər ehtiva edən müşahidədir. Həddindən artıq göstərici nəticəyə mane ola bilər, buna görə də bunun qarşısını almaq lazımdır.
- Multikollinearlıq: Müstəqil dəyişənlər digər dəyişənlərə nisbətən bir-biri ilə yüksək korrelyasiyaya malikdirsə, bu şərt Multikollinearlıq adlanır. O, verilənlər bazasında olmamalıdır, çünki o, ən çox təsir edən dəyişəni sıralayarkən problem yaradır.
- Yetərsiz uyğunlaşdırma və həddindən artıq uyğunlaşdırma: Əgər alqoritmimiz təlim verilənlər toplusu ilə yaxşı işləyirsə, lakin test verilənlər toplusu ilə yaxşı deyilsə, belə problem overfitting adlanır. Və əgər alqoritmimiz hətta təlim verilənlər bazası ilə də yaxşı performans göstərmirsə, belə bir problemə uyğunsuzluq deyilir.

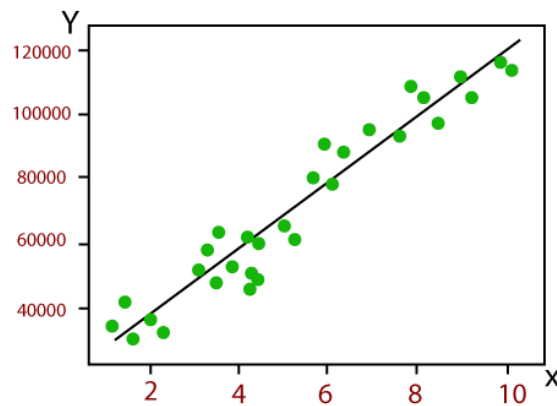
Niyə reqressiya analizindən istifadə edilir? Yuxarıda qeyd edildiyi kimi, reqressiya təhlili davamlı dəyişənin proqnozlaşdırılmasına kömək edir. Real dünyada müxtəlif ssenarilər var ki, burada hava şəraiti, satış proqnozu, marketinq meylləri və s. kimi gələcək proqnozlara ehtiyacımız var, belə halda proqnozları daha dəqiq edə bilən bəzi texnologiyaya ehtiyacımız var. Beləliklə, belə bir vəziyyət üçün statistik bir üsul olan və maşın öyrənməsi və məlumat elmində istifadə olunan Reqressiya təhlilinə ehtiyacımız var. Aşağıda reqressiya analizindən istifadə etmək üçün bəzi digər səbəblər verilmişdir:

- Reqressiya hədəf və müstəqil dəyişən arasındakı əlaqəni təxmin edir.
- Məlumatlarda meylləri tapmaq üçün istifadə olunur.
- Real, davamlı dəyərləri proqnozlaşdırmağa kömək edir.
- Reqressiyanı həyata keçirməklə biz ən vacib amili, ən az vacib amili və hər bir faktorun digər amillərə necə təsir etdiyini əminliklə müəyyən edə bilərik.

Reqressiyanın növləri: Məlumat elmində və maşın öyrənməsində istifadə olunan reqressiyaların müxtəlif növləri var. Hər növün müxtəlif ssenarilərdə öz əhəmiyyəti var, lakin əsasda bütün reqressiya metodları müstəqil dəyişənin asılı dəyişənlərə təsirini təhlil edir. Bəzi mühüm reqressiya növlərinə xətti reqressiya, logistik

reqressiya, polinom reqressiyası, dəstək vektor reqressiyası, qərar ağacının reqressiyası, təsadüfi meşə reqressiyası, ridge reqressiyası, lasso reqressiyası misal göstərmək olar.

Xətti reqressiya: Xətti reqressiya nəzarət edilən öyrənmə texnikasına daxil olan və reqressiya problemlərini həll etmək üçün istifadə edilən ən sadə maşın öyrənmə alqoritmlərindən biridir. Müstəqil dəyişənlərin köməyi ilə davamlı asılı dəyişəni proqnozlaşdırmaq üçün istifadə olunur. Xətti reqressiyanın məqsədi davamlı asılı dəyişən üçün çıxışı dəqiq proqnozlaşdırabilən ən yaxşı uyğun xətti tapmaqdır. Proqnoz üçün tək müstəqil dəyişən istifadə olunursa, o, sadə xətti reqressiya adlanır və ikidən çox müstəqil dəyişən varsa, belə reqressiya çox xətti reqressiya adlanır. Ən yaxşı uyğunluq xəttini tapmaqla, alqoritm asılı dəyişən və müstəqil dəyişən arasında əlaqə qurur və əlaqə xətti xarakterli olmalıdır. Xətti reqressiyanın nəticəsi yalnız qiymət, yaş, əmək haqqı və s. kimi davamlı dəyərlər olmalıdır. Asılı dəyişən və müstəqil dəyişən arasındakı əlaqə aşağıdakı şəkildə göstərilə bilər:



Şək 2.1 Xətti reqressiya

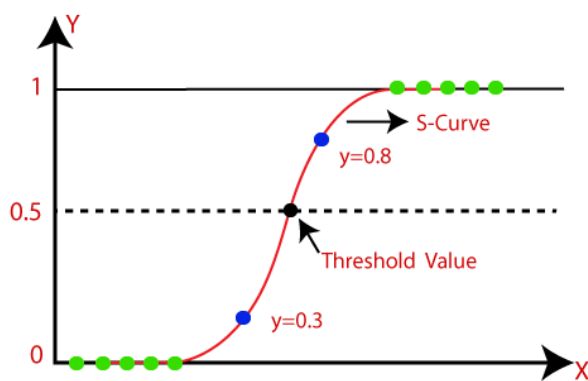
Yuxarıdakı şəkildəki asılı dəyişən Y oxu (maaş) və müstəqil dəyişən x oxu (təcrübə) üzərindədir. Reqressiya xətti belə yazıla bilər:

$$(2.1)$$

Burada β_0 və β_1 əmsallar, ϵ isə xəta terminidir.

Logistik reqressiya: Logistik reqressiya nəzarət edilən öyrənmə üsulları altında olan ən məşhur maşın öyrənmə alqoritmlərindən biridir. Təsnifat və reqressiya

problemləri üçün istifadə edilə bilər, lakin əsasən təsnifat problemləri üçün istifadə olunur. Müstəqil dəyişənlərin köməyi ilə kateqoriyalı asılı dəyişəni proqnozlaşdırmaq üçün logistik reqressiyadan istifadə olunur. Logistik reqressiya probleminin nəticəsi yalnız 0 ilə 1 arasında ola bilər. Logistik reqressiya iki sinif arasındakı ehtimalların tələb olunduğu yerlərdə istifadə edilə bilər. Məsələn, 0 və ya 1, doğru və ya yalan və s. Logistik reqressiya maksimum ehtimalın qiymətləndirilməsi konsepsiyasına əsaslanır. Bu qiymətləndirməyə görə, müşahidə edilən məlumatlar ən çox ehtimal olunan olmalıdır. Logistik reqressiyada biz daxilolmaların çəkili cəmini 0 və 1 arasında dəyərləri xəritələyə bilən aktivləşdirmə funksiyasından keçirik. Belə aktivləşdirmə funksiyası sigmoid funksiyası, əldə edilən əyri isə sigmoid əyri və ya S əyrisi adlanır. Aşağıdakı şəkli nəzərdən keçirin:



Şək.2.2 Logistik reqressiya

Logistik reqressiya tənliyi:

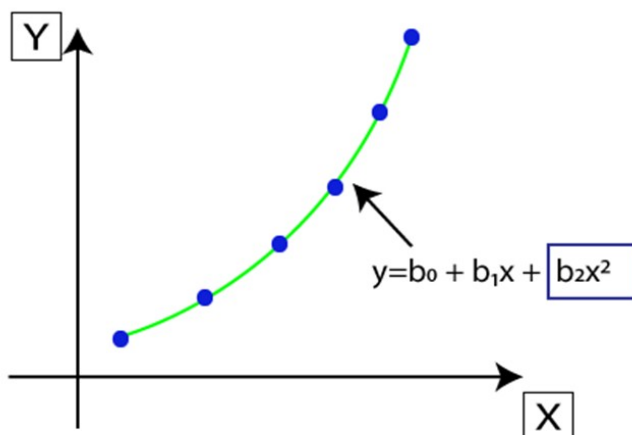
$$(2.2)$$

Xətti reqressiya və logistik reqressiya arasında fərqlər bunlardır:

- Xətti reqressiya, verilmiş müstəqil dəyişənlər toplusundan istifadə edərək davamlı asılı dəyişəni proqnozlaşdırmaq üçün istifadə olunur. Logistik reqressiya isə, verilmiş müstəqil dəyişənlər toplusundan istifadə edərək kateqoriyalı asılı dəyişəni proqnozlaşdırmaq üçün istifadə olunur.
- Xətti reqressiya reqressiya problemini həll etmək üçün istifadə olunur. Təsnifat problemlərinin həlli üçün isə logistik reqressiyadan istifadə olunur.

- Xətti reqressiyada biz davamlı dəyişənlərin qiymətini proqnozlaşdırırıq. Lakin logistik reqressiyada biz kateqoriyalı dəyişənlərin dəyərlərini proqnozlaşdırırıq.
- Xətti reqressiyada çıxışı asanlıqla proqnozlaşdırıla biləcəyimiz ən yaxşı uyğun xətti tapırıq. Logistik reqressiyada biz nümunələri təsnif edə biləcəyimiz S əyrisini tapırıq.
- Xətti reqressiyada dəqiqliyin qiymətləndirilməsi üçün ən kiçik kvadratın qiymətləndirilməsi metodundan istifadə edilir. Logistik reqressiyada isə dəqiqliyin qiymətləndirilməsi üçün maksimum ehtimalın qiymətləndirilməsi metodundan istifadə edilir.
- Xətti reqressiya üçün çıxış qiymət, yaş və s. kimi davamlı dəyər olmalıdır. Logistik Reqressiyanın çıxışı 0 və ya 1, Bəli və ya Xeyr və s. kimi kateqorik dəyər olmalıdır.
- Xətti reqressiyada asılı dəyişən ilə müstəqil dəyişən arasındakı əlaqə xətti olmalıdır. Logistik reqressiyada asılı və müstəqil dəyişən arasında xətti əlaqənin olması tələb olunmur.
- Xətti reqressiyada müstəqil dəyişənlər arasında kollinearlıq ola bilər. Logistik reqressiyada müstəqil dəyişən arasında kollinearlıq olmamalıdır.

Polinom reqressiyası: Polinom reqressiyası xətti modeldən istifadə edərək qeyri-xətti verilənlər toplusunu modelləşdirən reqressiya növüdür. O, çoxsaylı xətti reqressiyaya bənzəyir, lakin o, x dəyəri ilə y -nin müvafiq şərti qiymətləri arasında qeyri-xətti əyriyə uyğundur. Tutaq ki, qeyri-xətti şəkildə mövcud olan məlumat nöqtələrindən ibarət bir verilənlər bazası var, buna görə də belə vəziyyətdə xətti reqressiya bu məlumat nöqtələrinə ən yaxşı uyğun gəlməyəcək. Belə məlumat nöqtələrini əhatə etmək üçün bizə polinom reqressiya lazımdır. Polinom reqressiyasında ilkin xüsusiyyətlər verilmiş dərəcənin çoxhədli xüsusiyyətlərinə çevrilir və sonra xətti modeldən istifadə edərək modelləşdirilir. Bu o deməkdir ki, məlumat nöqtələri polinom xəttindən istifadə etməklə ən yaxşı şəkildə uyğunlaşdırılır.



Şək 2.3 Polinom reqressiya

Çoxhədli reqressiya tənliyi də $Y = b_0 + b_1x$ xətti reqressiya tənliyi mənasını verən xətti reqressiya tənliyindən alınmışdır,

(2.3)

polinom reqressiya tənliyinə çevrilir. Burada Y proqnozlaşdırılan yeni hədəf məhsul, b_0, b_1, \dots, b_n reqressiya əmsallarıdır. x bizim müstəqil yeni giriş dəyişənimizdir. Model hələ də xəttidir, çünki əmsallar hələ də kvadratla xəttidir

Dəstək vektor reqressiyası (SVR): SVR reqressiya, eləcə də təsnifat problemləri üçün istifadə edilə bilən nəzarət edilən öyrənmə alqoritmidir. Beləliklə, onu reqressiya problemləri üçün istifadə etsək, o zaman dəstək vektor reqressiyası adlanır. Bu alqoritm davamlı dəyişənlər üçün işləyən reqressiya alqoritmidir. Aşağıda dəstək vektor reqressiyasında istifadə olunan bəzi açar sözlər verilmişdir:

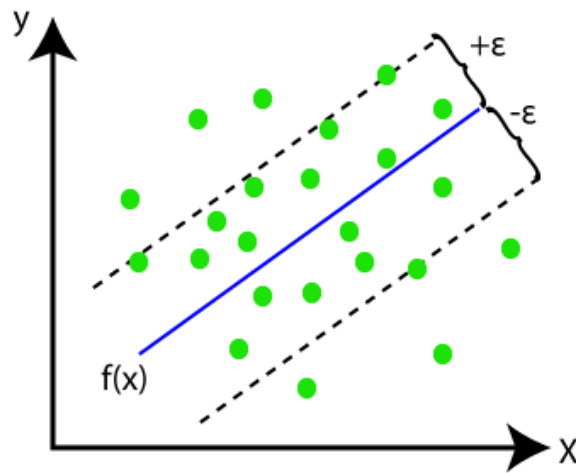
Kernel: Bu, aşağı ölçülü məlumatı daha yüksək ölçülü məlumatlara uyğunlaşdırmaq üçün istifadə olunan bir funksiyadır.

Hiperplan: Ümumiyyətlə, SVM, iki sinif arasında bir ayırma xəttidir, lakin SVR-də davamlı dəyişənləri proqnozlaşdırmağa kömək edən və məlumat nöqtələrinin əksəriyyətini əhatə edən bir xəttidir.

Sərhəd xətti: Sərhəd xətləri məlumat nöqtələri üçün kənar yaradan hipertəpədən ayrı iki xəttidir.

Dəstək vektorları: Dəstək vektorları hipertəpəyə ən yaxın və əks sinif olan məlumat nöqtələridir.

SVR-də biz həmişə maksimum ölçüyə malik hipertəpəni müəyyən etməyə çalışırıq ki, məlumat nöqtələrinin maksimum sayı həmin marjada əhatə olunsun. SVR-nin əsas məqsədi sərhəd xətləri daxilində maksimum məlumat nöqtələrini nəzərə almaqdır və hiperplanda (ən yaxşı uyğun xətt) maksimum sayda məlumat nöqtəsi olmalıdır.



Şək 2.4 Dəstək Vektor reqressiya

Burada mavi xətt hiperplan adlanır, digər iki xətt isə sərhəd xətləri kimi tanınır.

Qərar ağacı həm təsnifat, həm də reqressiya problemləri üçün istifadə oluna bilən nəzarət olunan öyrənmə texnikasıdır, lakin əsasən təsnifat problemlərinin həlli üçün üstünlük verilir. Qərar ağacı, ağac strukturlu təsnifatçıdır, burada daxili qovşaqlar məlumat dəstinin xüsusiyyətlərini, budaqlar qərar qaydalarını və hər bir yarpaq qovşağının nəticəsini təmsil edir.

Qərar ağacında iki qovşaq var, bunlar qərar düyünü və yarpaq düyünüdür. Qərar qovşaqları hər hansı bir qərar qəbul etmək üçün istifadə olunur və çoxlu budaqlara malikdir, yarpaq qovşaqları isə bu qərarların nəticəsidir və başqa filialları ehtiva etmir.

Qərarlar və ya sınaq verilmiş verilənlər toplusunun xüsusiyyətləri əsasında həyata keçirilir. Bu, verilmiş şərtlər əsasında problemin, qərarın bütün mümkün həll yollarını əldə etmək üçün qrafik təsvirdir. O, qərar ağacı adlanır, çünki o, ağaca bənzəyir, kök düyünlə başlayır və sonrakı budaqlarda genişlənir və ağaca bənzər bir quruluş yaradır. Ağac yaratmaq üçün biz təsnifat və reqressiya ağacı alqoritmini ifadə edən CART alqoritmindən istifadə edirik. Qərar ağacı sadəcə sual verir və cavaba əsasən ağacı daha da alt ağaclara ayırır. Qərar ağaclarından istifadə olunmasının səbəbi: maşın öyrənməsində müxtəlif alqoritmlər var, ona görə də verilmiş verilənlər toplusu və problem üçün ən yaxşı alqoritmin seçilməsi maşın öyrənmə modelini yaradarkən yadda saxlamalı olan əsas məqamdır.

Aşağıda qərar ağacından istifadə etməyin iki səbəbi var:

- Qərar Ağacları qərar qəbul edərkən adətən insanın düşünmə qabiliyyətini təqlid edir, buna görə də onu başa düşmək asandır.
- Qərar ağacının arxasındakı məntiq asanlıqla başa düşülə bilər, çünki o, ağaca bənzər bir quruluş göstərir.

Qərar ağacı terminologiyasına aşağıdakılar aiddir:

- Kök qovşağı: Kök düyünü qərar ağacının başladığı yerdir. Daha sonra iki və ya daha çox homojen dəstlərə bölünən bütün məlumat dəstini təmsil edir.
- Yarpaq qovşağı: Yarpaq düyünləri son çıxış qovşağıdır və yarpaq düyünü əldə etdikdən sonra ağac daha çox ayrılı bilməz.
- Parçalanma: Parçalanma qərar qovşağının və ya kök qovşağının verilmiş şərtlərə uyğun olaraq alt qovşaqlara bölünməsi prosesidir.
- Budaq, alt ağac: Ağacın parçalanması ilə əmələ gələn ağac.
- Budama: Budama ağacdən arzuolunmaz budaqların çıxarılması prosesidir.
- Valideyn, uşaq qovşağı: Ağacın kök düyünü ana qovşaq, digər qovşaqlar isə uşaq qovşaqları adlanır.

Qərar ağacı alqoritm necə işləyir? Qərar ağacında, verilmiş verilənlər toplusunun sinifini proqnozlaşdırmaq üçün alqoritm ağacın kök qovşağından başlayır. Bu

alqoritm kök atributunun qiymətlərini rekord yəni, real verilənlər toplusu atributu ilə müqayisə edir və müqayisəyə əsasən budaqı izləyir və növbəti düyünə keçir. Növbəti qovşaq üçün alqoritm yenidən atribut dəyərini digər alt qovşaqlarla müqayisə edir və daha da irəliləyir. Ağacın yarpaq düyününə çatana qədər prosesi davam etdirir. Tam prosesi aşağıdakı alqoritmdən istifadə etməklə daha yaxşı başa düşmək olar:

Addım 1: Ağac kök düyün ilə başlanılır, tam məlumat dəsti S ilə işarə olunur.

Addım 2: Atribut seçim ölçüsündən istifadə edərək verilənlər bazasında ən yaxşı atributu tapılır.

Addım 3: S ən yaxşı atributlar üçün mümkün dəyərləri ehtiva edən alt çoxluqlara bölünür.

Addım 4: Ən yaxşı atributu ehtiva edən qərar ağacı düyünü yaradılır.

Addım 5: 3-cü addımda yaradılmış verilənlər toplusunun alt çoxluqlarından istifadə edərək rekursiv olaraq yeni qərar ağacları yaradılır. Düyünləri daha çox təsnif edə bilməyəcəyiniz və son nodu yarpaq qovşağı adlandırma bilməyəcəyiniz mərhələyə çatana qədər bu prosesi davam etdirilir.

Atribut seçimi tədbirləri: Qərar ağacını həyata keçirərkən əsas məsələ kök qovşaq və alt qovşaqlar üçün ən yaxşı atributun necə seçiləcəyi ilə bağlıdır. Beləliklə, bu cür problemləri həll etmək üçün atribut seçimi ölçüsü və ya ASM adlanan bir texnika var. Bu ölçü ilə ağacın düyünləri üçün ən yaxşı atributu asanlıqla seçə bilərik. ASM üçün iki məşhur texnika var, bunlar məlumat qazancı və Gini indeksidir.

1. Məlumat qazancı: Məlumat qazancı, atribut əsasında verilənlər toplusunun seqmentləşdirilməsindən sonra entropiyada dəyişikliklərin ölçülməsidir. O, funksiyanın bir sinif haqqında bizə nə qədər məlumat verdiyini hesablayır. Məlumat qazancının dəyərinə görə, qovşağı parçalayırıq və qərar ağacı qururuq. Qərar ağacı alqoritmı həmişə məlumat qazancının dəyərini maksimuma çatdırmağa çalışır və ən yüksək məlumat qazancına malik olan atribut əvvəlcə bölünür. Məlumat qazancı bu düsturla hesablanma bilər:

(2.4)

Burada entropiya müəyyən bir atributdakı çatışmayan dəyəri ölçmək üçün bir metrikdir. Məlumatda təsadüfiliyi müəyyən edir. Entropiya aşağıdakı kimi hesablanabilir:

(2.5)

Burada S - nümunələrin ümumi sayı, $P(bəli)$ - bəli ehtimalı, $P(yox)$ - yox ehtimalıdır.

2. Gini İndeksi- CART yəni, təsnifat və reqressiya ağacı alqoritmində qərar ağacı yaradarkən istifadə olunan yanlış və ya doğruluğun ölçüsüdür. Yüksək Gini indeksi ilə müqayisədə aşağı Gini indeksi olan bir atribut üstünlük verilməlidir. O, yalnız ikili bölünmələr yaradır və CART alqoritmi ikili bölünmələr yaratmaq üçün Gini indeksindən istifadə edir. Gini indeksi aşağıdakı düsturla hesablanabilir:

(2.6)

Qərar ağacının üstünlükləri:

- İnsanın real həyatda hər hansı bir qərar qəbul edərkən izlədiyi eyni prosesi izlədiyi üçün başa düşmək asandır.
- Qərarla bağlı problemlərin həlli üçün çox faydalı ola bilər.
- Problemin bütün mümkün nəticələrini düşünməyə kömək edir.
- Digər alqoritmlərlə müqayisədə məlumatların təmizlənməsi tələbi daha azdır.

Qərar ağacının çatışmazlıqları:

- Qərar ağacında çoxlu təbəqə var, bu da onu mürəkkəb edir.
- Təsadüfi meşə alqoritmi ilə həll edilə bilən həddindən artıq uyğunlaşma problemi ola bilər.
- Daha çox sinif etikətləri üçün qərar ağacının hesablama mürəkkəbliyi arta bilər.

Təsadüfi meşə, nəzarət edilən öyrənmə texnikasına aid olan məşhur maşın öyrənmə alqoritmidir. Maşın öyrənməsində həm təsnifat, həm də reqressiya

problemləri üçün istifadə edilə bilər. Mürəkkəb problemi həll etmək və modelin performansını yaxşılaşdırmaq üçün bir neçə təsnifatçının birləşdirilməsi prosesi olan ansambl öyrənmə konsepsiyasına əsaslanır.

Təsadüfi meşə alqoritmi: Adından da göründüyü kimi, "Təsadüfi meşə, verilmiş məlumat dəstinin müxtəlif alt dəstləri üzrə bir sıra qərar ağaclarını ehtiva edən və həmin məlumat dəstinin proqnozlaşdırma dəqiqliyini yaxşılaşdırmaq üçün orta hesabla götürən təsnifatçıdır." Bir qərar ağacına güvənmək əvəzinə, təsadüfi meşə hər bir ağacdən proqnoz alır və proqnozların səs çoxluğuna əsaslanaraq yekun nəticəni proqnozlaşdırır. Təsadüfi meşədə daha çox ağac olması daha yüksək dəqiqliyə gətirib çıxarır və həddindən artıq uyğunlaşma probleminin qarşısını alır. Təsadüfi meşə verilənlər toplusunun sinifini proqnozlaşdırmaq üçün bir neçə ağacı birləşdirdiyinə görə, bəzi qərar ağacları düzgün çıxışı proqnozlaşdırır, digərləri isə olmaya bilər. Ancaq birlikdə, bütün ağaclar düzgün çıxışı proqnozlaşdırır. Buna görə də, aşağıda daha yaxşı Təsadüfi meşə təsnifatı üçün iki fərziyyə mövcuddur. Birincisi, verilənlər dəstinin xüsusiyyət dəyişməsində bəzi faktiki dəyərlər olmalıdır ki, təsnifatlandırıcı təxmin edilən nəticədən çox dəqiq nəticələri proqnozlaşdırır. İkincisi, hər bir ağacdən gələn proqnozlar çox aşağı korrelyasiyaya malik olmalıdır.

Aşağıda təsadüfi meşə alqoritmindən niyə istifadə etməli olduğumuzu izah edən bəzi məqamlar verilmişdir:

- Digər alqoritmlərlə müqayisədə daha az məşq vaxtı tələb olunur.
- O, hətta səmərəli işlədiyi böyük verilənlər toplusu daxil olmaqla yüksək dəqiqliklə çıxışı proqnozlaşdırır.
- Təsadüfi meşə, həmçinin məlumatların böyük bir hissəsi əskik olduqda dəqiqliyi qoruya bilər.

Təsadüfi meşə iki mərhələdə işləyir, birinci növbədə n qərar ağacını birləşdirərək təsadüfi meşə yaratmaq, ikincisi isə birinci mərhələdə yaradılmış hər bir ağac üçün proqnozlar verməkdir. İş prosesi aşağıdakı addımlarda izah edilə bilər:

Addım 1: Təlim dəstindən təsadüfi K məlumat nöqtələri seçilir.

Addım 2: Seçilmiş məlumat nöqtələri (Alt qruplar) ilə əlaqəli qərar ağacları qurulur.

Addım 3: Qurmaq istədiyiniz qərar ağacları üçün N rəqəmi seçilir.

Addım 4: Addım 1 və 2-i təkrarlanır.

Addım 5: Yeni məlumat nöqtələri üçün hər bir qərar ağacının proqnozlarını tapılır və yeni məlumat nöqtələrini səs çoxluğunu qazanan kateqoriyaya təyin edilir.

Təsadüfi meşənin üstünlükləri:

- Təsadüfi meşə həm təsnifat, həm də reqressiya tapşırıqlarını yerinə yetirməyə qadirdir.
- O, yüksək ölçüyə malik böyük verilənlər toplusunu idarə etməyə qadirdir.
- Bu, modelin dəqiqliyini artırır və həddindən artıq uyğunlaşma probleminin qarşısını alır.

Təsadüfi meşənin çatışmazlıqları: Təsadüfi meşə həm təsnifat, həm də reqressiya tapşırıqları üçün istifadə oluna bilsə də, reqressiya tapşırıqları üçün daha uyğun deyil.

Ridge reqressiyası: Ridge reqressiyası xətti reqressiyanın ən möhkəm versiyalarından biridir ki, burada daha yaxşı uzunmüddətli proqnozlar əldə etmək üçün az miqdarda qərəz tətbiq edilir. Modelə əlavə edilən qərəz miqdarı Ridge reqressiya penalty kimi tanınır. Bu cəza müddətini lambda ilə hər bir fərdi xüsusiyyətin kvadrat çəkisinə vurmaqla hesablaya bilərik.

Silsilənin reqressiyası üçün tənlik belə olacaq:

$$(2.7)$$

Müstəqil dəyişənlər arasında yüksək kollinearlıq olarsa, ümumi xətti və ya çoxhədli reqressiya uğursuz olacaq, ona görə də belə problemləri həll etmək üçün Ridge reqressiyasından istifadə etmək olar. Ridge reqressiyası, modelin mürəkkəbliyini azaltmaq üçün istifadə edilən bir nizamlama üsuludur. Buna L2

nizamlanması da deyilir. Nümunələrdən daha çox parametrimiz varsa, problemləri həll etməyə kömək edir.

Lasso Reqressiyası: Lasso reqressiyası modelin mürəkkəbliyini azaltmaq üçün başqa bir nizamlama üsuludur. Bu, Ridge reqressiyasına bənzəyir, istisna olmaqla, cəza müddəti çəkilərin kvadratı əvəzinə yalnız mütləq çəkiləri ehtiva edir. Mütləq dəyərlər qəbul etdiyinə görə, o, yamacı 0-a qədər kiçilə bilər, Ridge reqressiyası isə onu yalnız 0-a yaxın azalda bilər. Buna L1 nizamlanması da deyilir. Lasso reqressiyası üçün tənlik belə olacaq:

(2.8)

K-ən yaxın qonşu alqoritmi- nəzarət edilən öyrənmə texnikasına əsaslanan ən sadə maşın öyrənmə alqoritmlərindən biridir. Bu alqoritmi yeni hal, data və mövcud hallar arasında oxşarlığı qəbul edir və yeni işi mövcud kateqoriyalara ən çox oxşar olan kateqoriyaya qoyur. K ən yaxın qonşu alqoritmi bütün mövcud məlumatları saxlayır və oxşarlıq əsasında yeni məlumat nöqtəsini təsnif edir. Bu o deməkdir ki, yeni məlumatlar görünəndə bu alqoritmədən istifadə etməklə asanlıqla quyu dəstləri kateqoriyasına təsnif edilə bilər. K ən yaxın qonşu alqoritmi həm reqressiya, həm də təsnifat üçün istifadə oluna bilər, lakin daha çox təsnifat məsələlərində istifadə olunur. K ən yaxın qonşu alqoritmi qeyri-parametrik bir alqoritmədir, yəni əsas məlumatlar üzərində heç bir fərziyyə irəli sürmür. Bu alqoritm həm də tənbəl öyrənən alqoritmi adlanır, çünki o, təlim toplusundan dərhal öyrənmir, əvəzinə verilənlər toplusunu saxlayır və təsnifat zamanı verilənlər toplusunda hərəkət edir. Təlim mərhələsində K ən yaxın qonşu alqoritmi yalnız verilənlər bazasını saxlayır və o, yeni məlumatlar əldə etdikdən sonra həmin məlumatları yeni verilənlərə çox oxşar olan kateqoriyaya təsnif edir. Tutaq ki, iki kateqoriya var, məsələn, A kateqoriyası və B kateqoriyası və bizdə yeni məlumat nöqtəsi x_1 var, ona görə də bu məlumat nöqtəsi bu kateqoriyalardan hansında yerləşəcək. Bu tip problemi həll etmək üçün bizə K ən yaxın qonşu alqoritmi lazımdır. Bu alqoritm köməyi ilə biz müəyyən verilənlər toplusunun kateqoriyasını və ya sinfini asanlıqla müəyyən edə bilərik. K ən yaxın qonşu alqoritmının işini aşağıdakı alqoritm əsasında izah etmək olar:

Addım 1: Qonşuların K sayını seçilir

Addım 2: K qonşuların Evklid məsafəsini hesablanılır

Addım 3: Hesablanmış Evklid məsafəsinə görə ən yaxın K qonşunu götürülür

Addım 4: Bu k qonşular arasında hər bir kateqoriyadakı məlumat nöqtələrinin sayı hesablanır

Addım 5: Yeni məlumat nöqtələrini qonşunun sayının maksimum olduğu kateqoriyaya təyin edilir

Məlumat nöqtələri arasındakı Evklid məsafəsi iki nöqtə arasındakı məsafədir. Bunu aşağıdakı kimi hesablamaq olar:

(2.9)

K ən yaxın qonşu alqoritmində K dəyərini necə seçmək olar? K-NN alqoritmində K dəyərini seçərkən yadda saxlamaq lazım olan bəzi məqamlar var: “K” üçün ən yaxşı dəyəri müəyyən etməyin xüsusi yolu yoxdur, ona görə də onlardan ən yaxşısını tapmaq üçün bəzi dəyərləri sınımalıyıq. K üçün ən çox seçilən dəyər 5-dir. $K=1$ və ya $K=2$ kimi K üçün çox aşağı dəyər səs-küylü ola bilər və modeldə kənar göstəricilərin təsirinə səbəb ola bilər. K üçün böyük dəyərlər yaxşıdır, lakin bəzi çətinliklərlə üzləşə bilər.

K ən yaxın qonşu alqoritminin üstünlükləri:

- Onun həyata keçirilməsi sadədir.
- Səs-küylü məşq məlumatlarına qarşı davamlıdır
- Təlim məlumatları böyükdürsə, daha effektiv ola bilər.

K ən yaxın qonşu alqoritminin çatışmazlıqları:

- Həmişə bir müddət mürəkkəb ola bilən K-nin dəyərini müəyyən etmək lazımdır.
- Bütün təlim nümunələri üçün məlumat nöqtələri arasındakı məsafənin hesablanması səbəbindən hesablama dəyəri yüksəkdir.

Maşın öyrənməsi alqoritmlərinin qiymətləndirilməsi, yəni onlar vasitəsilə qurulan modelin nə dərəcədə dəqiq olmasını müəyyən etmək üçün bir çox göstəricilərdən istifadə olunur. Bu dissertasiya işində isə, qurulan modelin alqoritmlərinin qiymətləndirilməsi üçün r-kvadrat, mütləq orta xəta və mütləq kvadratik xətdən istifadə olunmuşdur.

R-kvadrat- reqressiya analizində istifadə olunan statistik ölçüdür və müstəqil dəyişənlərin asılı dəyişəni nə qədər yaxşı izah etdiyini göstərir. R-kvadrat dəyəri reqressiya modelinin məlumat dəstindəki fərqi neçə faizini izah etdiyini ifadə edir. Yəni daha yüksək R-kvadrat dəyəri o deməkdir ki, model verilənlər dəstindəki variasiyanın daha böyük hissəsini izah edir və buna görə də model verilənlərə nə qədər yaxşı uyğun gəlir. R-kvadrat dəyəri adətən 0 ilə 1 arasında qiymət alır. Ən pis vəziyyətdə R-kvadrat dəyəri 0-dır, ən yaxşı halda R-kvadrat dəyəri 1-ə yaxınlaşır. Mənfi dəyərlər də mümkündür, lakin bunlar adətən model məlumatlara uyğun gəlmədikdə baş verir. R-kvadrat reqressiya analizində modelin keyfiyyətini qiymətləndirmək üçün istifadə olunur. Lakin təkcə R-kvadrat dəyərindən modelin dəqiqliyi haqqında tam təsəvvür əldə etmək mümkün deyil. Digər qiymətləndirmə meyarları da nəzərə alınmalıdır.

Mütləq orta xəta (MAE)- modelin keyfiyyətini və düzgünlüyünü qiymətləndirmək üçün istifadə olunan göstəricidir. Mütləq orta xətanı tapmaq üçün modelin proqnozlaşdırdığı qiymət və real qiymətlərin fərqinin orta qiyməti hesablanır.

Mütləq kvadratik xəta (MSE)- statistika və maşın öyrənməsi sahələrində istifadə olunan səhv ölçüsüdür. Xüsusilə reqressiya problemlərində modelin proqnozlarının real dəyərlərdən nə qədər uzaq olduğunu ölçmək üçün geniş istifadə olunur. MSE proqnozlaşdırılan dəyərlər və faktiki dəyərlər arasındakı fərqlərin kvadratlarının ortasını hesablayır. Daha dəqiq desək, hər bir məlumat nöqtəsi üçün proqnozlaşdırılan dəyər ilə faktiki dəyər arasındakı fərq kvadratlaşdırılır və bu dəyərlər orta hesablanır. MSE-nin üstünlüklərindən biri mənfi səhvlərin kvadratlaşdırma xətalrı səbəbindən müsbət səhvlərlə kompensasiya edilməsidir.

MSE dəyəri nə qədər kiçik olsa, modelin proqnozları həqiqi dəyərlərə bir o qədər yaxındır.

2.5 Maşın öyrənməsi həyat dövrü

Maşın öyrənməsi kompüter sistemlərinə açıq şəkildə proqramlaşdırılmadan avtomatik öyrənmə qabiliyyəti verir. Əsas suallardan biri isə maşın öyrənmə sisteminin necə işlədiyini ilə bağlıdır. Beləliklə, bu proses maşın öyrənməsinin həyat dövründən istifadə edərək təsvir edilə bilər. Maşın öyrənməsinin həyat dövrü səmərəli maşın öyrənmə layihəsini qurmaq üçün dövrü bir prosesdir. Həyat dövrünün əsas məqsədi isə problemin və ya layihənin həllini tapmaqdır.

Maşın öyrənməsinin həyat dövrü aşağıda verilmiş yeddi əsas addımı əhatə edir:

- Məlumatların toplanması
- Məlumatların hazırlanması
- Məlumatın təmizlənməsi
- Məlumatları təhlil edilməsi
- Modelin öyrədilməsi
- Modelin sınaqması
- Modelin tətbiqi

Tam prosesdə ən vacib şey problemi başa düşmək və problemin məqsədini bilməkdir. Buna görə də, həyat dövrünə başlamazdan əvvəl problemi başa düşmək vacibdir, çünki yaxşı nəticə problemin daha yaxşı başa düşülməsindən asılıdır. Tam həyat dövrü prosesində bir problemi həll etmək üçün "model" adlı maşın öyrənmə sistemi yaradırıq və bu model "təlim" təmin edilərək yaradılır. Lakin bir modeli hazırlamaq üçün bizə məlumat lazımdır, buna görə də həyat dövrü məlumatların toplanması ilə başlayır.

1. Məlumatların toplanması: Məlumatların toplanması maşın öyrənməsinin həyat dövrünün ilk addımıdır. Bu addımın əsas məqsədi məlumatla bağlı bütün problemləri müəyyən etmək və əldə etməkdir. Bu addımda müxtəlif məlumat mənbələrini

müəyyən etməliyik, çünki məlumatlar fayllar, verilənlər bazası, internet və ya mobil cihazlar kimi müxtəlif mənbələrdən toplanı bilər. Bu, həyat dövrünün ən vacib mərhələlərindən biridir. Toplanan məlumatların kəmiyyət və keyfiyyəti çıxışın səmərəliliyini müəyyən edəcəkdir. Çünki, məlumat nə qədər çox olarsa, proqnoz bir o qədər dəqiq olacaqdır. Bu addıma müxtəlif məlumat mənbələrini müəyyənləşdirmək, məlumat toplamaq, müxtəlif mənbələrdən əldə edilən məlumatları birləşdirmək kimi tapşırıqlar daxildir. Bu tapşırığı yerinə yetirməklə, verilənlər toplusu kimi də adlandırılan əlaqəli məlumat dəstini əldə edirik hansı ki, həmin məlumatlar bundan sonrakı addımlarda istifadə olunacaq.

2. Məlumatların hazırlanması: Məlumatları topladıqdan sonra onu növbəti addımlar üçün hazırlamalıyıq. Məlumatların hazırlanması məlumatlarımızı uyğun yerə qoyduğumuz və maşın öyrənmə təlimimizdə istifadə etmək üçün hazırladığımız bir addımdır. Bu addımda, ilk növbədə, biz bütün məlumatları bir yerə yığıq və sonra məlumatların sıralanmasını təsadüfiləşdiririk. Bu addımı daha iki prosesə bölmək olar:

- Məlumat kəşfiyyəti: O, işləməli olduğumuz məlumatların təbiətini anlamaq üçün istifadə olunur. Biz məlumatların xüsusiyyətlərini, formatını və keyfiyyətini başa düşməliyik. Məlumatların daha yaxşı başa düşülməsi öz növbəsində effektiv nəticəyə gətirib çıxarır. Burada korrelyasiya, ümumi tendensiyalar və kənar göstəricilər tapırıq.
- Məlumatın əvvəlcədən işlənməsi: Növbəti addım verilənlərin təhlili üçün əvvəlcədən işlənməsidir.

3. Məlumatların təmizlənməsi: Məlumat mübahisəsi xam məlumatların təmizlənməsi və istifadə edilə bilən formata çevrilməsi prosesidir. Bu, məlumatların təmizlənməsi, istifadə ediləcək dəyişənin seçilməsi və növbəti mərhələdə təhlil üçün daha uyğun olması üçün məlumatların düzgün formata çevrilməsi prosesidir. Bu mərhələ tam prosesin ən vacib addımlarından biridir. Keyfiyyət problemlərini həll etmək üçün məlumatların təmizlənməsi tələb olunur. Topladığımız məlumatların həmişə bizim istifadəmizdə olması vacib deyil, çünki bəzi məlumatlar faydalı olmaya

bilər. Real dünya tətbiqlərində toplanmış məlumatların müxtəlif problemləri ola bilər, o cümlədən: çatışmayan dəyərlər, dublikat məlumatlar, yanlış data, səs-küy (noise) və s. Beləliklə, məlumatları təmizləmək üçün müxtəlif filtrləmə üsullarından istifadə edilir. Yuxarıda göstərilən problemlərin aşkar edilməsi və aradan qaldırılması məcburidir, çünki bu, nəticənin keyfiyyətinə mənfi təsir göstərə bilər.

4. Məlumatların Təhlili: Bu mərhələdə təmizlənmiş və hazırlanmış məlumatlar artıq təhlil mərhələsinə keçir. Bu addıma analitik üsulların seçilməsi, modellərin qurulması və nəticəni nəzərdən keçirilməsi daxildir. Bu addımın məqsədi müxtəlif analitik üsullardan istifadə edərək məlumatları təhlil etmək və nəticəni nəzərdən keçirmək üçün maşın öyrənmə modelini qurmaqdır. Mərhələ problemlərin növünün müəyyən edilməsi ilə başlayır. Burada təsnifat, reqressiya, klaster təhlili, assosiasiya və s. kimi maşın öyrənmə üsulları seçilir, daha sonra hazırlanmış məlumatlardan istifadə edərək modeli qurulur və sonda qurulan modeli qiymətləndirilir.

5. Modelin öyrədilməsi: İndi növbəti addım modeli öyrətməkdir, bu addımda problemin daha yaxşı nəticəsi üçün onun performansını yaxşılaşdırmaq üçün modelimizi öyrədilir. Biz müxtəlif maşın öyrənmə alqoritmlərindən istifadə edərək modeli öyrətmək üçün məlumat dəstlərindən istifadə edilir. Modelin müxtəlif nümunələri, qaydaları və xüsusiyyətləri anlaya bilməsi üçün təlim tələb olunur.

6. Test Modeli: Maşın öyrənmə modelimiz verilmiş verilənlər toplusunda öyrədildikdən sonra modeli sınaqdan keçirilir. Bu addımda biz modelimizin düzgünlüyünü ona test məlumat dəsti təqdim etməklə yoxlayırıq. Modelin sınaqdan keçirilməsi layihənin və ya problemin tələbinə uyğun olaraq modelin faiz dəqiqliyini müəyyən edir.

7. Maşın öyrənməsinin həyat dövrünün son addımı, modeli real dünya sistemində tətbiq edilməsidir. Əgər hazırlanmış model məqbul sürətlə tələbimizə uyğun dəqiq nəticə verirsə, biz modeli real sistemdə yerləşdiririk. Ancaq layihəni tətbiq etməzdən əvvəl, mövcud məlumatlardan istifadə edərək onun performansını yaxşılaşdırıb-

yaxşılaşmadığını yoxlayacağıq. Yerləşdirmə mərhələsi layihə üçün yekun hesabatın hazırlanmasına bənzəyir.

2.6 Maşın öyrənməsində istifadə olunan vasitələr

Maşın öyrənmə proqramlarını inkişaf etdirmək üçün bir sıra müxtəlif alətlər və kitabxanalar var. Maşın öyrənmə sahəsində tez-tez istifadə olunan bəzi alətlər bunlardır:

Python: Python maşın öyrənməsi və süni intellekt tətbiqləri üçün ən populyar proqramlaşdırma dillərindən biridir. Python açıq mənbədir və müxtəlif maşın öyrənmə alqoritmləri və alətlərini özündə birləşdirən böyük kitabxana ekosisteminə malikdir. Pythonun bir çox üstünlükləri var. Aşağıda bu üstünlüklərdən bəziləri qeyd edilmişdir:

- Sadə və aydın sintaksis: Python oxumaq və yazmaq asan olan sintaksisə malikdir. Sadə və başa düşülən strukturu ilə yeni başlayanlar üçün öyrənmək asandır.
- Çox məqsədli proqramlaşdırma: Python çoxməqsədli proqramlaşdırma dilidir. O, veb inkişafında, məlumatların təhlili, süni intellekt, informasiya təhlükəsizliyi, elmi hesablamalar və bir çox başqa sahələrdə istifadə edilə bilər.
- Geniş kitabxana dəstəyi: Python standart kitabxanaların böyük dəstinə malikdir. Bu kitabxanalarda müxtəlif tapşırıqlar üçün istifadə oluna bilən və Python-un müxtəlif sahələrdə istifadə edilməsinə imkan verən bir çox funksiya var.
- İcma dəstəyi: Python-un geniş və aktiv icması var. Bu icma sənədlər, yardım forumları, kitabxana inkişafı və başlanğıc resursları daxil olmaqla bir çox resurs təqdim edir.
- Obyekt yönümlü proqramlaşdırma : Python obyekt yönümlü proqramlaşdırmadan faydalanır. Bu, kodun daha modul, təkrar istifadə edilə bilən və saxlanıla bilən olmasına imkan verir.

- Sürətli prototipləmə: Python sürətli prototipləmə üçün ideal dil hesab olunur. Asan sintaksis və geniş kitabxana dəstəyi ideyaların tez bir zamanda prototiplənməsini və sınaqdan keçirilməsini asanlaşdırır.
- Açıq mənbə və ödənişsiz olması: Python açıq mənbəli layihədir və pulsuz istifadə edilə bilər. Bu, Python-u hər kəsin əldə edə və istifadə edə biləcəyi bir dil edir. Əlavə olaraq Python müxtəlif platformalarda işləyə bilər, bu da kodunuzun müxtəlif əməliyyat sistemlərində asanlıqla işləməsinə imkan verir. Pythonun maşın öyrənməsində istifadə olunan bir çox kitabxanası mövcuddur. Onlardan bəziləri aşağıdakılardır:

Scikit-learn: Scikit-learn, Python-a əsaslanan açıq mənbəli maşın öyrənmə kitabxanasıdır. Buraya təsnifat, reqressiya, klasterləşdirmə, ölçülərin azaldılması, model seçimi və modelin qiymətləndirilməsi kimi bir çox maşın öyrənməsi alqoritmləri daxildir.

TensorFlow: Google tərəfindən hazırlanmış TensorFlow, dərin öyrənmə modellərini qurmaq və öyrətmək üçün istifadə edilən məşhur açıq mənbəli maşın öyrənmə kitabxanasıdır. O, qrafik əsaslı hesablama çərçivəsini təklif edir və çoxqatlı neyron şəbəkələrinin, konvolyusiya neyron şəbəkələrinin və təkrarlanan neyron şəbəkələrinin asan yaradılmasına imkan verir.

Keras: Keras TensorFlow, Microsoft Cognitive Toolkit, Theano kimi müxtəlif arxa hissələrdə işləyə bilən yüksək səviyyəli dərin öyrənmə kitabxanasıdır. Dərin öyrənmə modellərini asanlıqla və tez yaratmağa imkan verir.

PyTorch: PyTorch, Facebook tərəfindən hazırlanmış açıq mənbəli maşın öyrənmə kitabxanasıdır. O, dinamik hesablama qrafiklərindən istifadə edir və tədqiqat məqsədləri üçün dərin öyrənmə modellərinin qurulması üçün xüsusilə populyardır.

Pandas: Pandas verilənlərin təhlili və məlumatların manipulyasiyası üçün istifadə edilən Python əsaslı kitabxanadır. O, məlumat dəstlərini yükləmək, təmizləmək, çevirmək və təhlil etmək üçün güclü alətlər təklif edir.

NumPy: NumPy Python-da elmi hesablamalar aparmaq üçün əsas kitabxanadır. NumPy massivləri verilənlər üzərində vektor və matris əməliyyatlarını yerinə yetirmək üçün sürətli və səmərəli üsul təqdim edir.

Matplotlib və Seaborn: Matplotlib və Seaborn vizuallaşdırma üçün istifadə olunan Python kitabxanalarıdır. Onlar qrafiklərlə məlumatları vizuallaşdırmaq və təhlil etmək üçün müxtəlif alətlər təqdim edir.

III FƏSİL MODELİN QURULMASI MƏRHƏLƏLƏRİ

3.1 Datanın toplanılması və vizuallaşdırılması

Maşın öyrənməsinin həyat dövründə qeyd etdildiyi kimi modelin qurulması bir neçə mərhələdən ibarətdir. Onlardan birincisi mövzuya uyğun datanın toplanılmasıdır.

Datanın toplanılması: Modelin qurulmasında istifadə olunacaq data İnformasiya və telekommunikasiya texnologiyaları kafedrasının birinci kurs tələbələrin birinci smestr üzrə təhsil aldıkları fənlərlərdən olan qiymətlər , ümumi orta müvəffəqiyyət göstəricisi, üniveristetə qəbul balı kimi akademik göstəricilərdən, cins, yaş kimi demoqrafik göstəricilərdən, valideynlərinin təhsil səviyyəsi, ailənin gəlir səviyyəsi, yaşadığı ünvanın universitetə yaxınlıq məsafəsi kimi isə sosial-iqtisadi və dərslərin hazırlanmasına ayırdığı müddət, məşğulluq vəziyyəti, və digər məlumatlardan ibarətdir. Məlumatlar csv yəni vergüllə ayrılmış data formasında modelə əlavə olunacaq. Onuda qeyd edim ki, model Python proqramlaşdırma dilində və bu dili dəstəkləyən Visual Studio İDE-də yəni inteqrasiya olunmuş inkişaf mühiti olan proqram təminatında yazılmışdır.

Data daxil edilməmişdən qabaq modelin qurulması üçün lazımlı olan kitabxanalar daxil edilməlidir. Aşağıda kod vasitəsilə kitabxanalar daxil edilir:

```
# kitabxanaların daxil edilməsi
import pandas as pd
import numpy as np
```

✓ 50.1s

Şək 3.1 Kitabxanaların daxil edilməsi

Datanın daxil edilməsi üçün pandas kitabxanasının read_csv() funksiyasından istifadə edilir:

```
telebedata=pd.read_csv(r"C:\Users\Casper\Downloads\datauni3tt.csv")
telebedata
```

Şək 3.2 Datasetin daxil edilməsi

Datanın üzərində dəyişiklik edəcəyimiz zaman əvvəlki formasının olduğu kimi qalması üçün datasetin adını yeni bir adla `copy()` vasitəsilə əvəzləyirik.

```
df=telebedata.copy()
```

Şək3.3 Datasetin başqa adla saxlanması

Daxil edilən data 100-dən böyük olduqda bəzi sətirlər və sütunlar gizlədilər. Bütün sətir və sütunları tam şəkildə görmək üçün aşağıdakı kodu daxil edirik:

```
pd.set_option('display.max_rows',None)  
pd.set_option('display.max_columns',None)
```

Şək 3.4 Datasetdə bütün məlumatların görünməsi üçün kod

Modelin qurulmasında istifadə olunan datalar haqqında ətraflı məlumat etmək lazımdır. Datanın hansı adda sütunlardan ibarət olduğunu bilmək üçün `columns` xüsusiyyətindən, neçə sətir və sütundan ibarət olduğunu bilmək üçün `shape` xüsusiyyətlərindən istifadə olunur. Lakin ən yaxşı üsul `info()` funksiyasından istifadə etməkdir. `Info` bizə qeyd etdiimiz məlumatlarla yanaşı datadakı verilənlərin tipini, `null` yəni, boş qiymət olub olmadığını və yaddaş həcmi də göstərir.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 124 entries, 0 to 123
Data columns (total 23 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   telebeid                                                                124 non-null    int64
1   Fizika                                                                  123 non-null    float64
2   Informasiya texnologiyaları əsasları  124 non-null    int64
3   proqramlaşmanın əsası                                                  118 non-null    float64
4   Azərbaycan dilində isguzar                                             124 non-null    int64
5   xetti cebr ve analitik hendese                                         124 non-null    int64
6   xarici dil                                                                123 non-null    float64
7   Cins                                                                      124 non-null    object
8   Qebul bali                                                                124 non-null    float64
9   Odenis formasi                                                           124 non-null    object
10  İş                                                                        124 non-null    object
11  Fealiyyet                                                                124 non-null    object
12  vaxt                                                                      124 non-null    object
13  əlavəkurs                                                                124 non-null    object
14  Dərsdav                                                                  124 non-null    object
15  Valideyn                                                                124 non-null    object
16  gəlir                                                                    124 non-null    object
17  məsafəsi                                                                124 non-null    object
18  Sağlamlıq                                                                124 non-null    object
19  ixtisas                                                                124 non-null    object
```

Şək 3.5 Datalar haqqında ümumi məlumat

Bu cədvəldən aydın olduğu kimi, data 23 sütün və 124 sətirdən ibarətdir və hər sütünün hansı tipdə olduğu, null qiymətə sahib olub-olmaması haqqında məlumatlar öz əksini tapmışdır.

Datanın minimum, maksimum, say, orta qiymət və s.kimi statistiki məlumatlarını görmək üçün isə describe() funksiyasından istifadə edilir.

```
df.describe()
```

✓ 0.3s Python

	telebeid	Fizika	Informasiya texnologiyaları əsasları	proqramlaşmanın əsası	Azərbaycan dilində isguzar	xetti cebr və analitik hendese	xarici dil	Qebul bali	ÜOMG
count	124.000000	123.000000	124.000000	118.000000	124.000000	124.000000	123.000000	124.000000	124.000000
mean	62.500000	76.439024	81.612903	68.533898	81.733871	73.475806	82.878049	400.232258	76.790323
std	35.939764	10.289376	10.849832	11.493352	11.970906	9.908905	10.344544	39.746597	9.779915
min	1.000000	51.000000	54.000000	51.000000	51.000000	51.000000	51.000000	311.600000	47.000000
25%	31.750000	69.000000	74.000000	59.250000	74.500000	66.000000	77.000000	381.025000	71.750000
50%	62.500000	75.000000	82.000000	68.500000	84.000000	74.000000	84.000000	392.650000	78.000000
75%	93.250000	84.000000	90.000000	76.000000	91.000000	81.000000	91.000000	417.925000	84.000000
max	124.000000	98.000000	100.000000	97.000000	100.000000	95.000000	100.000000	542.800000	96.000000

Şək 3.5 Datalar haqqında statistiki məlumatlar

Hər dəfə bütöv datanın görünməsinə ehtiyac yoxdursa `head()` funksiyası ilə yalnız ilk 5 sətirinin görünməsindən istifadə edə bilərik.

Dataların vizuallaşdırılması: Datasetdəki verilənləri vizual olaraq təsvir edərək dataset haqqında daha asan məlumat sahibi ola bilərik. Vizuallaşdırma üçün `matplotlib` və `seaborn` kitabxanalarını daxil edirik.

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
# matplotlibin subplot funksiyası ilə iki qrafikin yaradılması
```

```
f,ax=plt.subplots(1,2,figsize=(8,6))
```

```
# seaborn kitabxanasının countplot funksiyası ilə qrafikin qurulması
```

```
sns.countplot(x=df['Cins'],data=df,palette='bright',ax=ax[0],saturation=0.95)
```

```
for container in ax[0].containers:
```

```
    ax[0].bar_label(container,color='black',size=15)
```

```
# X və y-in adını Cins və say olaraq və ölçüsünü təyin etmək
```

```
ax[0].set_xlabel('Cins', fontsize=14)
```

```
ax[0].set_ylabel('Say', fontsize=14)
```

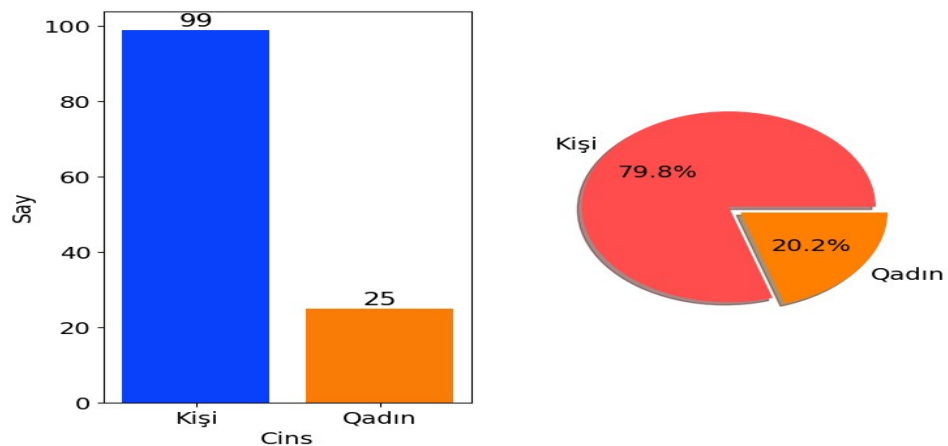
```
ax[0].tick_params(labels=14)
```

```
# pie qrafikin qurulması
```



```
plt.pie(x=df
['Cins'].value_counts(),labels=['Kişi','Qadın'],explode=[0,0.1],autopct='%1.1f%%',shadow=True,colors=['#ff4d4d','#ff8000'], textprops={'fontsize': 14})
# Qrafikin təsvir edilməsi
plt.show()
```

Beləliklə aşağıdakı qrafik alınır:



Şək 3.6 Datasetdəki cins sütününün təsviri

Qrafiklər vasitəsilə datalardakı outlier yəni kənar qiymətləri asanlıqla müşahidə etmək olur. Bunun üçün adətən histqram qurulur.

```
# histqramın qurulması
```

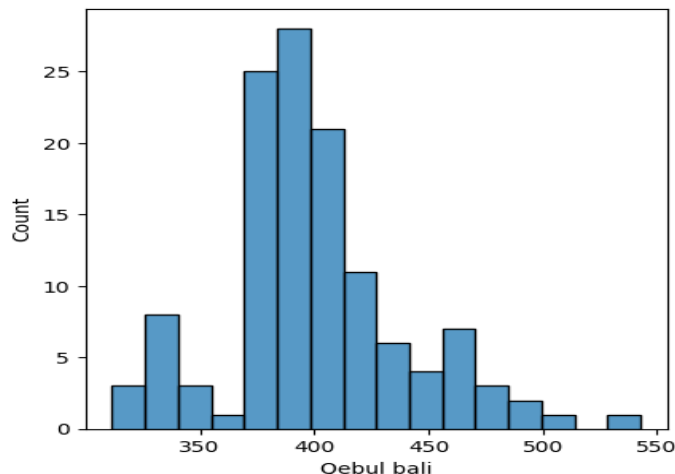
```
plt.figure(figsize = (16,5))
```

```
#sns.distplot(data['writing score'])
```

```
plt.subplot(1, 3, 1)
```

```
sns.histplot(df['Qebul balı'])
```

```
plt.show()
```



Şək 3.7 Qəbul balı sütünunu təsvir edən histoqram

3.2 Məlumatların emalı , təmizlənməsi

Null qiymətlərin müəyyən olunması və onların aradan qaldırılması:

Datasetdəki verilənlər haqqında ümumi məlumata sahib olduqdan sonra ikinci mərhələ datanın təmizlənməsi mərhələsidir. Datanın təmizlənməsinə datada null yəni boş qiymətlərin olub olmadığını yoxlamaq, dublikat məlumatları araşdırmaq, kateqorik məlumatların aşkarlanması və çevrilməsi və s. aiddir.

Datasetdə null qiymətlərin olması maşın öyrənmə alqoritmlərinin düzgün proqnozlaşdırma etməsində problem yarada bilər. Bunun qarşısını almaq üçün null qiymətləri model öyrədilməsindən əvvəl aşkarlamalıyıq. Datasetdə null qiymətlərinin olub- olmamasını `isnull()` funksiyası ilə və sayını isə `sum()` funksiyası ilə müəyyən edə bilərik.

```
df.isnull().sum()
✓ 0.1s
```

telebeid	0
Fizika	1
Informasiya texnologiyaları əsasları	0
proqramlaşmanın əsası	6
Azərbaycan dilində isguzar	0
xetti cebr ve analitik hendese	0
xarici dil	1
Cins	0

Şək 3.8 Datasetdə null qiymətlərinin yoxlanılması

Göründüyü kimi bəzi sütünlarda null qiyməti var. Daha düzgün model qurmaq üçün bu qiymətləri aradan qaldırmaq lazımdır. Bunun üçün müxtəlif yanaşmalar mövcuddur. Ən sadə həll dropna() funksiyasından istifadə etməklə null qiymətlərinin olduğu sətiri görə görməkdir. Lakin bu üsul bir null qiymətə sətirdəki digər məlumatların itməsinə səbəb olur. Buna görə də praktik olaraq bu üsul xüsusilə kiçik verilənlər bazası üçün əlverişli hesab olunmur. Digər bir üsul isə fillna() funksiyası ilə boş xanaları fərqli dəyərlərlə doldurmaq üsuludur. Fillna() funksiyasının iki əsas metodu var bfill və ffill. Bfill metodunda boş xanalar özündən sonrakı dolu xananın ortalaması qiymətini boş xanaya əlavə edir, ffill metodun isə özündən əvvəlki xananın qiymətini boş xanaya əlavə edir. Bu halda ən sadə üsul isə fillna() ilə xanaya sıfır qiyməti daxil etməkdir

```
df=df.fillna(0)
df.isnull().sum()
✓ 0.3s
```

telebeid	0
Fizika	0
Informasiya texnologiyaları əsasları	0
proqramlaşmanın əsası	0
Azərbaycan dilində isguzar	0
xetti cebr ve analitik hendese	0
xarici dil	0
Cins	0
Qebul bali	0

Şəx 3.9 Null qiymətlərin aradan qaldırılması

Datasetdə duplicat məlumatların olub olmadığını yoxlamaq üçün isə duplicated() funksiyasından istifadə edirik.

```
df.duplicated().sum()
✓ 0.4s
```

0

Şəx 3.10 datasetdə dublikat dəyərin yoxlanılması

Kateqorik məlumatların təyin olunması və kodlaşdırılması: Verilənlərin emalı mərhələsində əsas proseslərdən biri isə kateqorik məlumatların aşkarlanmasıdır. Kateqorik məlumatlar simvol tipli məlumatlardır və bu məlumatlar maşın öyrənmə alqoritmləri tərəfindən işləyə bilmirlər. Buna görə ilk növbədə həmin dataları müəyyənləşdirmək daha sonra isə onları numerik yəni rəqəm tipli datalara çevirmək tələb olunur.

```
#numerical and categorical columns sütunları müəyyən etmək
numeric_features = [feature for feature in df.columns if df[feature].dtype != "object"]
categorical_features = [feature for feature in df.columns if df[feature].dtype == "object"]

print("Datasetdə {} numerical sütun var: {}".format(len(numeric_features), numeric_features))
print("Datasetdə {} categorical sütun var: {}".format(len(categorical_features), categorical_features))
```

✓ 0.0s

Datasetdə 9 numerical sütun var: ['telebeid', 'Fizika', 'Informasiya texnologiyaları əsasları', 'proqram']
 Datasetdə 14 categorical sütun var: ['Cins', 'Odenis forması', 'İş', 'Fealiyyət', 'vaxt', 'əlavəkurs', '']

Şək 3.11 Kateqorik və rəqəm tipli sütunların müəyyənləşdirilməsi

Kateqorik məlumatların olduğu sütunları müəyən etdikdən sonra bu verilənləri numerik məlumatlara kodlaşdırmaq lazımdır. Bunun üçün scikit-learn kitabxanasının One-Hot Encoder və ya Label encoder kimi kodlaşdırma usullarından istifadə edilir.

```
# label encoder daxil edilmesi
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
for col in df:
    if type(df[col][0]) is str:
        df[col] = label_encoder.fit_transform(df[col])
        df[col].replace(inplace=True)
df.head()
```

Şək 3.12 Kateqorik dataların rəqəm tipli datalara çevrilməsi

Yuxarıdakı kod vasitəsilə bütün kateqorik məlumatlara sahib olan sütunlar rəqəm tipli datalara çevrilir.

3.3 Xüsusiyyət mühəndisliyi

Data ilə bağlı bütün təmizləmə mərhələlərini yerinə yetirdikdən sonra modelin qurulmasında növbəti addım olan xüsusiyyətlərin seçilməsini həyata keçiririk. Beləki maşın öyrədilməsi alqoritmləri vasitəsilə qurulan modeldə ilk növbədə hədəf-target sutun və features- xüsusiyyətlər adlanan təyin olmalıdır. Burada y modelin proqnozlaşdıracağı hədəf sütunu adlanır. X isə hədəf sütunun proqnozlaşdırılması üçün hədəf sutuna təsir edən sutunlarından ibarətdir. Tələbə müvəffəqiyyətinin proqnozlaşdırılması üçün qurulan model verilmiş sütunlar içərisində ümumi orta müvəffəqiyyət göstəricisini proqnozlaşdıracaq. Bunun üçün y yeni hədəf sütunu olaraq ÜOMG təyin olunur.

Datasetin telebeid sütunu istisna olmaqla digər sütunları tələbə müvəffəqiyyətinə təsir edən sütunlar olduğu üçün X yeni xüsusiyyət sütun olaraq təyin edilir.

```
X = df.drop(columns="ÜOMG",axis=1)
y = df["ÜOMG"]
```

```
X=df.drop(columns="telebeid",axis=1)
X.head()
```

Şək.3.13 X və y sütunlarının təyin olunması

Eyni zamanda telebeid telebe müvəffəqiyyətinin proqnozlaşdırılmasında təsir edən amil olmadığına görə bu sütunu drop() funksiyası ilə silirik.

3.4 Verilənlərin train və test datalarına ayrılması

Modelin qurulmasından əvvəlki addımlardan biridə verilənlərin train və test datalarına ayrılmasıdır. Bu proses ona görə önəmlidir ki, biz modeli yalnız bir

datasetdəki verilənlərlə öyrətsək, o zaman bu model yalnız həmin datalar üçün doğru nəticələr verəcək. Bu zaman modeli yeni datalar üzərində tətbiq etsək, model qeyri-dəqiq nəticələr əldə edəcək. Bu problemin qarşısını almaq üçün verilənlərin train və test datalarına bölünməsi ehtiyacı yaranır. Çünki modelin praktiki dəyəri onun yeni məlumatlar üzrə proqnozlar verməsi ilə müəyyənləşir. Başqa sözlə hazırlanmış modelin performansını yoxlayarkən onun qurulmasında iştirakı olmayan məlumatlardan istifadə edirik. Verilənlərin train və test mərhələlərinə bölünməsinin məqsədində məhz bu fikirlə bağlıdır. Yəni modelin praktiki performansını yoxlamaq üçün bəzi dataları modelin qurulması mərhələsindən xaric edirik və daha sonra həmin verilənlərdən modelin dəqiqliyini əvvəllər görmədiyi verilənlər üzərində yoxlamaq üçün istifadə edirik.

Verilənlər scikit-learn kitabxanasının `test_train_split()` funksiyası ilə test və train datalarına ayrılır. Train datalar modelin öyrədilməsi üçün istifadə olunur. Test datalar vasitəsi ilə isə modelin praktiki performansı yoxlanılır.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2,random_state=42)
X_train.shape, X_test.shape
```

✓ 12.0s

((99, 22), (25, 22))

Şək3.14 Dasetin train və test verilənlərinə ayrılması

Nəticədən görüldüyü kimi 99 sətir verilən modelin öyrədilməsi, 22 sətir data modelin dəqiqliyinin yoxlanılması üçün ayrılıb. Adətən datalar 80:20 və ya 70:30 nisbətində train və test verilənlərinə ayrılırlar. Burada `random_state` metodu istifadə olunur. Bunun səbəbi isə modelin təsadüfi yolla verilənləri train və test datalarına ayırmasıdır və `random_state` metodunun istifadə edilməsinin məqsədi hər dəfə model run olunduqda yəni işlədikdə modelin eyni dataları seçməsinə təmin edilməsidir.

3.5 Modelin öyrədilməsi

Modelin qurulması üçün əsas məsələ həll edilməli olan problemə uyğun olaraq maşın öyrənməsi alqoritminin seçilməsidir. İndi isə ikinci fəsildə qeyd olunan

reqresiyya alqoritmlərindən istifadə etməklə modeli öyrədək və model vasitəsilə proqnozlar verək.

Sadə formada modelin öyrədilməsi və modelin proqnozz verilməsi mərhələsini izah etmək üçün ilk olaraq qərar vermə alqoritmindən istifadə edək.

Bunun üçün scikit-learn kitabxanasından `DecisionTreeRegressor()` modelə daxil edirik. Daha sonra alqoritm vasitəsilə model train məlumatlarından istifadə edilərək öyrədilir. Bunun üçün `fit()` funksiyasından istifadə edilir.

```
from sklearn.tree import DecisionTreeRegressor
```

```
model=DecisionTreeRegressor(random_state=1)
model.fit(X_train,y_train)
```

Şək 3.15 Qərar ağacı daxil edilməsi və modelin öyrədilməsi

Ən sonda qurulmuş model vasitəsilə hədəf sutunu proqnozlaşdırılır. Bu zaman X yəni xüsusiyyət sutunlarının test üçün ayrılmış verilənlərindən istifadə edərək proqnozlaşdırmanı həyata keçirən `predict()` funksiyasından istifadə edilir və nəticələr massiv olaraq ekrana çıxır:

```
texmin=model.predict(X_test)
```

✓ 0.5s

```
array([72., 84., 87., 84., 78., 73., 95., 85., 66., 65., 86., 47., 65.,
       79., 85., 76., 81., 81., 77., 76., 72., 73., 75., 95., 73.])
```

Şək 3.16 Qərar ağacı alqoritmi ilə proqnozun verilməsi

Model artıq verilmiş qiymətlərdən nümunələr öyrənərək, sütunlar arasında qanunauyğunluqları müəyyən edərək test datalar üçün yeni qiymətlər proqnoz edir.

3.6 Modelin qiymətləndirilməsi

Modelin doğrulunu, yəni onun proqnoz etdiyi qiymətlərin real qiymətlərə nə dərəcədə yaxın olduğunu yoxlayaraq modelin performansını qiymətləndirə bilərik. Bunun üçün ikinci fəsilə qeyd edilmiş reqresiyya alqoritmlərinin keyfiyyətini yoxlayan orta mütəq xəta, orta kvadratik xəta, kvadrat kök xəta, R2 xəta kimi göstəricilərdən istifadə edilir. Bu model üçün orta mütləq xətanı hesablamaq üçün scikit-learn kitabxanasından mean-absolute-error() funksiyasından istifadə edirik.

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
mae=mean_absolute_error(y_test,texmin)
mse=mean_squared_error(y_test,texmin)
r2=r2_score(y_test,texmin)
print('Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mse))
print("- Mean Absolute Error: {:.4f}".format(mae))
print("- R2 score: {:.4f}".format(r2))
```

Modelin performansinin deqiqlik gostericileri:

- Mean Squared Error: 4.3600
- Mean Absolute Error: 0.6800
- R2 score: 0.9445

Şək 3.17 Qərar ağacı modelinin qiymətləndirilməsi

Modelin qurulmasında bir çox alqoritmdən istifadə olunması daha məqsədə uyğundur. Çünki, sonda müxtəlif alqoritmlər tərəfindən qurulan model üçün göstəricilərdən istifadə edərək qiymətləndirilən, ən az xəta vermiş alqoritm daha uyğun hesab olunacaq. Yuxarıda verilən prosessləri daha yığcam şəkildə təsvir etmək üçün istifadə olunan reqresiyya alqoritmlərinin hamısını eyni vaxtda modelə daxil edək:


```
# alqoritmlərin və göstəricilərin daxil edilməsi

from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor,AdaBoostRegressor
from sklearn.svm import SVR
from sklearn.linear_model import LinearRegression,Lasso
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
```

Şəil 3.18 Digər alqoritmlərin daxil edilməsi

Yuxarıda göstərilən addımlarla bu alqoritmlər üçündə model qurub qiymətləndirək.

Təsadüfi Meşə alqoritmi ilə modelin qurulması və qiymətləndirilməsi:

```
modelRF=RandomForestRegressor()
modelRF.fit(X_train,y_train)
texminRF=modelRF.predict(X_test)
texminRF
```

✓ 0.5s

```
array([71.89, 83.82, 87.01, 83.92, 77.9 , 72.9 , 93.95, 84.97, 67.59,
       63.47, 85.97, 54.74, 64.54, 79.02, 84.93, 75.97, 81.05, 81. ,
       76.93, 75.94, 71.8 , 72.94, 75.11, 93.14, 72.96])
```

Şək 3.19 Təsadüfi meşə modelinin nəticələri

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
maeRF=mean_absolute_error(y_test,texminRF)
mseRF=mean_squared_error(y_test,texminRF)
r2RF=r2_score(y_test,texminRF)
print('Təsadüfi meşə Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mseRF))
print("- Mean Absolute Error: {:.4f}".format(maeRF))
print("- R2 score: {:.4f}".format(r2RF))
```

Təsadüfi meşə Modelin performansinin deqiqlik gostericileri:

- Mean Squared Error: 0.6099
- Mean Absolute Error: 0.4336
- R2 score: 0.9922

Şək 3.20 Təsadüfi meşə modelinin qiymətləndirilməsi

Xətti regressiya ilə modelin qurulması və qiymətləndirilməsi:

```

modellLR=LinearRegression()
modellLR.fit(X_train,y_train)
texminLR=modellLR.predict(X_test)
texminLR

✓ 1.7s

array([[72., 84., 87., 84., 78., 73., 96., 85., 68., 64., 86., 57., 65.
       79., 85., 76., 81., 81., 77., 76., 72., 74., 75., 94., 74.]])

```

Şək 3.21 Xətti reqressiya modelinin nəticələri

```

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
maeLR=mean_absolute_error(y_test,texminLR)
mseLR=mean_squared_error(y_test,texminLR)
r2LR=r2_score(y_test,texminLR)
print('Xətti Reqresiyya Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mseLR))
print("- Mean Absolute Error: {:.4f}".format(maeLR))
print("- R2 score: {:.4f}".format(r2LR))

```

Xətti Reqresiyya Modelin performansinin deqiqlik gostericileri:
 - Mean Squared Error: 0.0000
 - Mean Absolute Error: 0.0000
 - R2 score: 1.0000

Şək 3.21 Xətti reqressiya modelinin qiymətləndirilməsi

K-qonşu reqressiya ilə modelin qurulması və qiymətləndirilməsi:

```

modelKN=KNeighborsRegressor()
modelKN.fit(X_train,y_train)
texminKN=modelKN.predict(X_test)
texminKN

1 ✓ 3.8s

array([[74.8, 84.8, 84.4, 85.2, 78. , 75. , 91.6, 83.8, 69.8, 66.8, 86.2,
       57.4, 68.4, 78.8, 85. , 78.4, 83.2, 81. , 76.2, 75.6, 71.8, 75.8,
       76. , 90.2, 74.2]])

```

Şək 3.22 K-qonşu modelinin nəticələri

```

maeKN=mean_absolute_error(y_test,texminKN)
mseKN=mean_squared_error(y_test,texminKN)
r2KN=r2_score(y_test,texminKN)
print('K Qonsu reqressor Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mseKN))
print("- Mean Absolute Error: {:.4f}".format(maeKN))
print("- R2 score: {:.4f}".format(r2KN))

```

K Qonsu reqressor Modelin performansinin deqiqlik gostericileri:
 - Mean Squared Error: 3.7808
 - Mean Absolute Error: 1.4640
 - R2 score: 0.9519

Şək 23 K-qonşu modelinin qiymətləndirilməsi

SVR alqoritmi ilə modelin qurulması və qiymətləndirilməsi:

```

models=SVR()
models.fit(X_train,y_train)
texminS=models.predict(X_test)
texminS

```

✓ 3.2s

```

array([77.60968384, 78.32586942, 78.43690866, 78.27256608, 77.93572861,
       77.62725662, 79.03908795, 78.35389387, 77.30151664, 77.05005181,
       78.45904738, 76.51893864, 77.1042285 , 77.92346926, 78.33891651,

```

Şək.3.24 Dəstək vektor maşını modelinin nəticələri

```

maeS=mean_absolute_error(y_test,texminS)
mseS=mean_squared_error(y_test,texminS)
r2S=r2_score(y_test,texminS)
print('SVR Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mseS))
print("- Mean Absolute Error: {:.4f}".format(maeS))
print("- R2 score: {:.4f}".format(r2S))

```

SVR Modelin performansinin deqiqlik gostericileri:

- Mean Squared Error: 69.0217
- Mean Absolute Error: 6.5511
- R2 score: 0.1219

Şək 3.25 SVR modelin qiymətləndirilməsi

AdaBoostRegressor alqoritmi ilə modelin qurulması və qiymətləndirilməsi:

```

modelA=AdaBoostRegressor()
modelA.fit(X_train,y_train)
texminA=modelA.predict(X_test)
texminA

```

✓ 4.4s

```

array([[70.75      , 84.      , 87.27272727, 84.      , 77.83333333,
        71.55555556, 93.875    , 85.      , 67.4      , 63.      ,
        86.60714286, 55.7      , 64.6      , 79.16666667, 85.      ,
        76.90909091, 81.5      , 81.5      , 77.18518519, 76.44444444,

```

Şək. 3.26 AdaBoostRegressor modelinin nəticələri

```

maeA=mean_absolute_error(y_test,texminA)
mseA=mean_squared_error(y_test,texminA)
r2A=r2_score(y_test,texminA)
print('AdaBoostRegressor Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mseA))
print("- Mean Absolute Error: {:.4f}".format(maeA))
print("- R2 score: {:.4f}".format(r2A))

```

AdaBoostRegressor Modelin performansinin deqiqlik gostericileri:

- Mean Squared Error: 0.6467
- Mean Absolute Error: 0.5615
- R2 score: 0.9918

Şək 27 AdaBoostRegressor modelinin qiymətləndirilməsi

Lasso algoritmi ilə modelin qurulması və qiymətləndirilməsi:

```

modell=Lasso()
modell.fit(X_train,y_train)
texminL=modell.predict(X_test)
texminL

```

✓ 4.7s

```

array([72.04115011, 83.87310875, 86.85248381, 83.92459075, 77.96905443,
       72.97552506, 95.79316286, 84.96080606, 68.12944097, 64.16779279,
       85.95996169, 57.06842597, 65.09569365, 78.89502963, 84.88670382,
       76.08210176, 80.92788891, 80.89573502, 77.04435625, 76.10135348,

```

Şək.3.28 Lasso modelinin nəticələri

```

maeL=mean_absolute_error(y_test,texminL)
mseL=mean_squared_error(y_test,texminL)
r2L=r2_score(y_test,texminL)
print('Lasso Modelin performansinin deqiqlik gostericileri:')
print("- Mean Squared Error: {:.4f}".format(mseL))
print("- Mean Absolute Error: {:.4f}".format(maeL))
print("- R2 score: {:.4f}".format(r2L))

```

✓ 0.0s

```

Lasso Modelin performansinin deqiqlik gostericileri:
- Mean Squared Error: 0.0105
- Mean Absolute Error: 0.0869
- R2 score: 0.9999

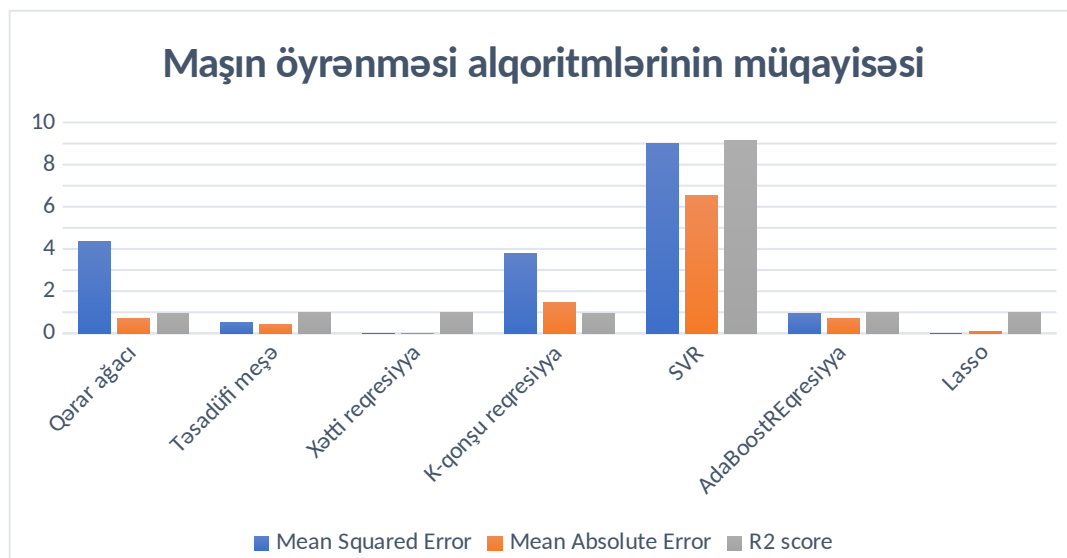
```

Şək. 3.29 Lasso modelinin qiymətləndirilməsi

Alqoritmlərin qiymətləndirmə göstəriciləri

	Mean Squared Error	Mean Absolute Error	R2 score
Qərar ağacı	4.36	0.68	0.9445
Təsadüfi meşə	0.5223	0.3968	0.9934
Xətti reqresiyya	0.002	0.001	1
K-qonşu reqresiyya	3.7808	1.464	0.9519
SVR	9.0217	6.5511	9.1219
AdaBoostReqresiyya	0.9323	0.7232	0.9881
Lasso	0.0105	0.0869	0.999

Bu cədvələ əsasən maşın öyrənmə alqoritmlərinin müqayisəsini aşağıdakı qrafiklə təsvir edək:



Şək 3.30 Alqoritmlərin göstəricilərinin müqayisəsi

Qrafikdən də göründüyü kimi, ən yaxşı performansə sahib alqoritmlər xətti reqressiya və Lasso alqoritmləri vasitəsilə qurulan modellərə aiddir. Xətti reqresiya ilə qurulmuş modeldə proqnozlaşdırılan ÜOMG görə hansı tələbələrin risk qrupunda olduğunu müəyyənləşdirək.

```

risk_grupu = df[df['ÜOMG'] < 65]['telebeid'].tolist()
print(risk_grupu)

```

✓ 0.0s

[5, 31, 52, 61, 63, 64, 85, 99, 100, 102, 105, 121]

Şək 3.31 Risk qrupunda olan tələbələrin müəyyən edilməsi

NƏTİCƏ

Sonda yazılan kod vasitəsilə proqnozlaşdırılmış ÜOMG içərisində akademik göstəricisi 65-dən kiçik olan tələbələrin id nömrələri müəyyən olunmuş və beləliklə id nömrələrə uyğun olaraq risk qrupunda olan tələbələrin siyahısı əldə olunmuşdur. Dissertasiyada seçilmiş məsələnin əhəmiyyəti araşdırılmış və həlli üçün model qurulmuşdur. Tələbənin akademik performansını və ya nailiyyətini proqnozlaşdırmaq tələbə akademik performans böhranını aradan qaldırmaq üçün vacib bir anlayışdır. Tələbələrin akademik müvəffəqiyyətini proqnozlaşdırmaq üçün modelin qurulmasında maşın öyrənməsinin müxtəlif reqressiya modellərindən istifadə edilmişdir. Modelin qurulmasının və tətbiqinin nəticəsi odur ki, o, pedaqoqlara bəzi sahələrdə çatışmazlıqları aşkar etməyə kömək edə bilər və zəif akademik göstəricilərə sahib tələbələrin erkən aşkarlanmasına imkan verməklə, müəllimlərə əsaslandırılmış qərarlar qəbul etmək və nəticədə tələbələrin akademik nailiyyətlərini və öyrənmə proseslərini təkmilləşdirməyə imkan verə bilər. Dissertasiya işi ilə bağlı təkliflər isə ondan ibarətdir ki, modelə gələcəkdə daha çox məlumatlar daxil edilərək daha yaxşı öyrədilə və beləliklə daha dəqiq nəticələr əldə edə bilərik. Çünki maşın öyrənməsi nə qədər çox nümunə ilə tanış olub həmin datalar əsasında öyrədilsə, qazanmış olduğu təcrübə o qədər çox olur və verilənlər arasında qanunauyğunluqları hətta gizli əlaqələri müəyyən edərək daha dəqiq nəticələr verə bilər. Eyni zamanda

model tələbələr haqqında məlumatların saxlanması üçün istifadə olunan sistemlərə inteqrasiya edilərək daha da təkmilləşdirilə bilər.

İSTİFADƏ EDİLMİŞ ƏDƏBİYYAT SİYAHISI

1. Tahiroğlu, K. (2021). Python ilə proqramlaşdırma
2. Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms.
<https://doi.org/10.1186/s40561-022-00192-z>.
3. Alyahyan, E., Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. <https://doi.org/10.1186/s41239-020-0177-7>
4. Alturki, S., Cohausz, L., Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: an empirical study in Germany.
<https://doi.org/10.1186/s40561-022-00220-y>
5. Kaur1, H., Tarandeep, K., Garg, K. (2022). A Prediction Model for Student Academic Performance Using Machine Learning.
<https://doi.org/10.31449/inf.v47i1.4297>
6. Garmaki, M., Boughzala, I., Wamba, S. (2016). The Effect Of Big Data Analytics Capability On Firm Performance.
<http://doi.org/10.1080/07421222.2018.1451955>

7. Admass, W. (2021). Review on Predicting Student Academic Performance using Data Mining Classification Algorithm.
<http://doi.org/10.1016/j.procs.2015.12.157>
8. Alturki, S., Cohausz, L., Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: an empirical study in Germany. <https://doi.org/10.1186/s40561-022-00220-y>.
9. Albreki, B., Nazar, Z. 2021. Review of Student' Performance Prediction Using Machine Learning Techniques.
<https://doi.org/10.3390/educsci11090552>
10. Manhaes, L., Manuel, S., Zimbrão, G. (2019). Towards Automatic Prediction of Student Performance in STEM Undergraduate Degree Programs.
<https://doi.org/10.1145/2695664.2695918>.
11. Marbouti, F., Madhavan, K. (2019). Models for early prediction at risk students in a course using standard based grading <https://doi.org/10.1016/j.compedu.2016.09.005>.
12. Santana, M.A., Costa, E.B., Fonseca, B., Rego, J. (2020). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. 247-256.
<https://doi.org/10.1016/j.chb.2017.01.047>.
13. Ihsan, A., Ashraf, Y. (2019). Students Performance Prediction Using KNN And Naïve Bayesian. <https://doi.org/10.1109/ICITECH.2017.8079967>
14. Krizanic, S. (2022). Educational Data Mining Using Cluster Analysis And Decision Tree Technique: A Case Study.
<http://doi.org/10.1177/1847979020908675>
15. Palacios, C., Leiva, V., Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. <http://dx.doi.org/10.3390/e23040485>

16. Rovira, S., Puertas, E., Igual, L. (2017). Data-driven system to predict academic grades and dropout. <https://doi.org/10.1371/journal.pone.0171207>
17. Muncie, T. (2020). Early Using Machine Learning Models to Predict Student Retention. <http://dx.doi.org/10.1038/s41598-023-32484-w>
18. Rao, G., Kumar, K. (2021). Students Performance Prediction in Online Courses Using Machine Learning Algorithms. <http://dx.doi.org/10.1109/IJCNN48605.2020.9207196>
19. Zeineddine, H., Braendle, U. (2021). Enhancing Prediction of Student Success: Automated Machine Learning Approach. <http://dx.doi.org/10.1016/j.compeleceng.2020.106903>
20. Berhanu, F. (2018). Students Performance Prediction Based on Their Academic Record. <http://dx.doi.org/10.5120/ijca2015907348>
21. Rodriguez, M. (2021). Prediction of University Students' Academic Achievement by Linear and Logistic Models. <http://dx.doi.org/10.1017/S1138741600004315>
22. Alamri, L., Almuslim, R., Aslam, N. (2020). Predicting Student Academic Performance Using Support Vector Machine and Random Forest. <http://dx.doi.org/10.1145/3446590.3446607>
23. Amjad, A. (2020). Prediction Of Students Performance Using Educational Data Mining. <http://dx.doi.org/10.14569/IJACSA.2016.070531>
24. Şahin, M., Keskin, S., Özgür, A., & Yurdugül, H. (2019). E-öğrenme ortamlarında öğrenen özelliklerine dayalı etkileşim profillerinin belirlenmesi. <http://dx.doi.org/10.17943/etku.297075>
25. Dalalyan, A. S., Hebiri, M., & Lederer, J. (2020). On the prediction performance of the lasso. 23(1), 552–581. <https://doi.org/10.3150/15-BEJ756>

26. Caldas, S. J., & Bankston, C. (2019). Effect of school population socioeconomic status on individual academic achievement. *90*(5), 269–277.
<https://doi.org/10.1080/00220671.1997.10544583>
27. Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *1*(3), 291–308. <https://doi.org/10.1257/aeri.20180633>
28. Dunlosky, J., Badali, S., Rivers, M. L., & Rawson, K. A. (2020). The role of effort in understanding educational achievement. *32*, 1163–1175.
<https://doi.org/10.1007/s10648-020-09577-3>
29. Huang, S., & Fang, N. (2019). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers and Education*, *61*(1), 133–145.
<https://doi.org/10.1016/j.compedu.2012.08.015>
30. Vandamme, J.-P., Meskens, N., & Superby, J.-F. (2021). Predicting academic performance by data mining methods. *Education Economics*, *15*(4), 405–419.
<https://doi.org/10.1080/09645290701409939>
31. Zabriskie, C., Yang, J., DeVore, S., & Stewart, J. (2019). Using machine learning to predict physics course outcomes. *Physical Review Physics Education Research*, *15*(2), 125-131
<https://doi.org/10.1103/PhysRevPhysEducRes.15.020120>
32. Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: A

machine-learning approach. *Higher Education*, 80(5), 875–894.

<https://doi.org/10.1007/s10734-020-00520-7>

33. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194.

<https://doi.org/10.1016/j.compedu.2017.05.007>