

UOT 004.65

TƏHSİLDƏ DATA MİNING METODLARININ TƏTBİQİ

F.T.Ağayev, N.İ.Bahadurzadə

Azərbaycan Respublikası Elm və Təhsil Nazirliyi

İnformasiya Texnologiyaları İnstitutu

E-mail: agayevinfo@gmail.com, bahadurzade99@mail.ru

Açar sözlər: Təhsil Data Miningi, Tələbə fəaliyyətinin proqnozlaşdırılması, Data Mining

Xülasə. Education Data Mining müasir dövrün inkişaf etməkdə olan sahələrindən biridir, təhsil sahəsindən tələbələr tərəfindən əldə edilən müxtəlif növ məlumatların araşdırılması üçün metodlar işləyib hazırlayır, həmçinin bu metodlar tələbələri və onların öyrəndikləri mühiti anlamağa kömək edir. Xüsusilə onlayn öyrənmə mühitində tələbə fəaliyyəti qiymətləndirilərkən təhsil sahəsində məlumatların öyrənilməsi mühüm rol oynayır. Education Data Mining təhsil məlumatlarını təhlil etmək üçün hesablama yanaşmalarından istifadə edir. Bu məqalənin məqsədi, Education Data Mining-də müxtəlif istifadəçi qruplarını, təhsil mühitlərinin növlərini və onların təqdim etdiyi məlumatları təkmilləşdirməkdən ibarətdir. Daha sonra təhsil mühitində data mining metodları vasitəsilə həll edilmiş tapşırıqlar və gələcək tədqiqatların ən perspektivli istiqamətlərindən bəziləri müzakirə edilir.

Giriş. E-təhsil data mining mühitində inkişaf edən bir sahədir. Bu rəqabətli dünyada təhsil mühiti həmçinin tələbə performansını araşdırmaq və təhlil etmək, məktəbi tərk etmənin qarşısını almaq, onların nəticələrini proqnozlaşdırmaq həm yaxşı, həm də akademik cəhətdən zəif tələbələrə diqqət yetirmək, fakültələr və müəllimlər üçün rəy, məlumatların vizuallaşdırılması və təlim prosesinin daha yaxşı qiymətləndirilməsi təhsilin keyfiyyəti üçün təkmilləşdirilir. Təhsil məlumatlarının öyrənilməsi bu təkmilləşdirmə üçün bir vasitədir. Müasir təhsil institutları öz strategiyaları və gələcək planları üçün məlumatların öyrənilməsinə ehtiyac duyurlar. Tələbənin fəaliyyəti fərdi, sosial, iqtisadi və digər ətraf mühit kimi müxtəlif amillərdən asılıdır. Yüksək səviyyəli təhsil müəssisələrinin səlahiyyətli orqanları yeni pedaqoji strategiyaların hazırlanmasına səbəb ola biləcək tələbələrin fəaliyyətdəki meyilləri və davranışları anlamaq üçün eksperimental nəticələrindən istifadə edirlər.

Educational Data Miningin (EDM) tədqiqat sahəsi təhsil mühitlərində məlumatların öyrənilməsinin texnika və üsullarının tətbiqinə yönəlmişdir. EDM, təhlili mümkün olmayan böyük təhsil məlumat növlərini araşdırmaq üçün maşın öyrənməsi, məlumatların öyrənilməsi və statistik metodların işləyib hazırlanması və tətbiqi ilə məşğul olur [1; s. 1-3]. Onların nəticələri tələbələrə öyrənmə performanslarını yaxşılaşdırmağa, fərdi tövsiyələr verməyə, tədris fəaliyyətini yaxşılaşdırmağa, təlimin effektivliyini qiymətləndirməyə, resursları və təhsil təklifini təşkil etməyə kömək edir. Digər tərəfdən, təhsildə İnternetdən istifadə e-öyrənmə və ya veb-əsaslı təhsil kimi tanınan yeni kontekst yaratdı ki, burada tədris-təlimin qarşılıqlı əlaqəsi haqqında böyük həcmdə məlumatı sonsuz şəkildə yaradır. Bütün bu məlumatlar təhsil məlumatlarının əsasını təmin edir. Educational Data Mining öyrənmə elminin ən yaxşı variantı və verilənlərin öyrənilməsinin bir qolu kimi qəbul edilə bilər [2,3,4]. Educational Data Mining istifadəçi qavrayışı, fəaliyyət və sınaq modelini yaradarkən faydalı ola bilər [5]. Data Mining o qədər populyarlıq qazandı ki, o, yeni aktuallığa çevrildi, çünki verilənlərin fərqli yanaşma formasını araşdırmaqda və onu funksional məlumatlara ixtisar etməkdə çox faydalıdır. EDM öyrənənləri daha yaxşı başa düşmək üçün bu məlumat bazalarından istifadə etməyə və bu nəzəriyyəni inkişaf etdirməyə çalışır. EDM son illərdə bütün dünyada müxtəlif və əlaqəli tədqiqat sahələrindən olan tədqiqatçılar üçün bir tədqiqat sahəsi olaraq ortaya çıxdı, məsələn:

– Offline təhsil üz-üzə təmas əsasında bilik və bacarıqları ötürür və həmçinin insanların necə öyrəndiyini psixoloji cəhətdən öyrənir.

– E-təhsil onlayn təlimat verir və bu sistemlər tərəfindən log fayllarında və verilənlər bazasında saxlanılan tələbə məlumatlarına Web Mining (WM) üsulları tətbiq edilir.

– İntellektual Repetitorluq (ITS - Intelligent Tutoring) və Adaptiv Təhsil Hipermedia Sistemi (Adaptive Educational Hypermedia System – AEHS) tədrisi hər bir konkret tələbənin ehtiyaclarına uyğunlaşdıraraq, sadəcə olaraq internetdə yerləşdirilən yanaşmaya alternativdir. Data Mining bu sistemlər tərəfindən toplanan məlumatlara, məsələn, log fayllarına, istifadəçi modellərinə və s. tətbiq edilmişdir.

EDM prosesi təhsil sistemlərindən gələn xam məlumatları potensial olaraq təhsil tədqiqatı və təcrübəsinə böyük təsir göstərə biləcək faydalı məlumatlara çevirir. Bu proses verilənlərin istehsalının biznes, genetik, tibb və s. kimi digər tətbiq sahələrindən çox da fərqlənmir, çünki o, ümumi məlumatların çıxarılması prosesi ilə eyni addımları izləyir: ilkin emal, verilənlərin əldə edilməsi və sonrakı emal.

Praktiki baxımdan EDM, təhsil sistemlərini qiymətləndirmək, təhsilin keyfiyyətinin bəzi aspektlərini potensial olaraq təkmilləşdirmək və daha effektiv təhsil yaratmaq məqsədilə tələbələrin istifadə məlumatlarına əsaslanan yeni bilikləri kəşf etməyə imkan verir. Bəzi oxşar ideyalar onlayn satışları artırma bilmək məqsədilə müştərilərin maraqlarını müəyyən etmək üçün verilənlərin əldə edilməsinin ilk və ən populyar tətbiqi olan e-ticarət sistemlərində uğurla tətbiq edilmişdir. Təhsildə bu istiqamətdə nisbətən az irəliləyiş olmasına baxmayaraq hazırda təhsil mühitində məlumatların öyrənilməsinin tətbiqinə maraq artmaqdadır. Bununla belə, DM-nin xüsusi olaraq təhsilə tətbiqini digər sahələrdə tətbiq olunduğundan fərqləndirən bəzi mühüm məsələlər var:

– Obyektiv. Hər bir tətbiq sahəsində verilənlərin əldə edilməsinin məqsədi fərqlidir. Məsələn, biznesdə əsas məqsəd pul məbləğləri, müştərilərin sayı və müştəri loyallığı ilə ölçülə bilən mənfəəti artırmaqdır. Lakin EDM həm öyrənmə prosesini təkmilləşdirmək, həm də tələbələrin öyrənməsinə rəhbərlik etmək kimi tətbiqi tədqiqat məqsədlərinə malikdir. Bu məqsədlərin kəmiyyətini müəyyən etmək bəzən çətin olur və onların xüsusi ölçmə üsulları tələb olunur.

– Data. Təhsil sistemində baza üçün çoxlu müxtəlif növ datalar mövcuddur. Bu verilənlər təhsil sahəsinə xasdır, belə ki, daxili semantik verilən, digər verilənlər ilə əlaqələr və mənalı iyerarxiyanın çoxsaylı səviyyələri var. Bəzi nümunələr, ITS və AEHS-də istifadə olunan, qrafik və ya iyerarxiya formatında müəyyən bir mövzunun anlayışları arasındakı əlaqələri təmsil edən domen modelidir (məsələn, kurs dərslərdə təşkil edilən bir neçə fəsildən ibarətdir və hər dərs bir neçə anlayışı ehtiva edir). Bundan əlavə, şagirdin və sistemin pedaqoji aspektlərini də nəzərə almaq lazımdır.

– Texnikalar. Təhsil məlumatlarının və problemlərin bəzi xüsusi xüsusiyyətləri var ki, bu da Data mining məsələsinin fərqli şəkildə həll olunmasını tələb edir. Ənənəvi DM üsullarının əksəriyyəti birbaşa tətbiq oluna bilsə də, digərləri konkret təhsil probleminə tətbiq oluna bilmir. Bundan əlavə, xüsusi məlumatların öyrənilməsi üsulları xüsusi təhsil problemləri üçün istifadə edilə bilər.

EDM müxtəlif istifadəçi qruplarını əhatə edir. Fərqli qruplar öz missiyalarına, baxışlarına və verilənlərin əldə edilməsindən istifadə məqsədlərinə uyğun olaraq təhsil məlumatlarına müxtəlif bucaqlardan baxırlar. Məsələn, EDM alqoritmləri ilə aşkar edilmiş biliklər yalnız müəllimlərə öz dərslərini idarə etmək, tələbələrinin öyrənmə proseslərini başa düşmək və öz tədris metodları üzərində düşünməkdə kömək etmək üçün deyil, həm də tələbənin vəziyyətlə bağlı fikirlərini dəstəkləmək və tələbələrə rəy bildirmək üçün istifadə edilə bilər.

Təhsildə Data Mining metodları. Data mining elementlərindən biri təhsil məlumatlarının istehsalıdır ki, onun əsas diqqəti tələbə məlumatlarından lazımi biliklərin əldə olunması üçün modellərin işlənilməsi və hazırlanmasıdır, onlardan istifadə etməklə tələbələrin akademik göstəriciləri artırıla bilər. Təhsil Data Mining prosesində xammal müxtəlif təhsil sistemlərindən qiymətli məlumatlara çevrilə bilər ki, bu da müəllimlər, tələbələr və onların valideynləri, təhsil tədqiqatçıları və təhsil proqram təminatı sisteminin tərtibatçıları tərəfindən istifadə edilə bilər. Təhsil məlumatlarının öyrənilməsi həm də mövcud təhsil sisteminin bir hissəsi olan və sistemin müxtəlif hissələri ilə müsbət qarşılıqlı əlaqə yarada bilən yeni model kimi də nəzərdən keçirilə bilər. Bu, ona son nəticədə tədrisin təkmilləşdirilməsi məqsədinə nail olmağa imkan verəcəkdir.

Educational Data Mining anlayışı təhsil sahəsindəki problemlərin həlli yollarını əldə etmək məqsədi ilə məlumatların analizi metodlarının təhsil məlumatlarının təhlilinə tətbiqidir. Təhsil tədqiqatlarının tətbiqində klasterləşdirmə, klassifikasiya, proqnozlaşdırma, maşın təlimi və son olaraq assosiasiya qaydalarının təhlili verilənlərin əldə edilməsi üçün ən geniş tanınan üsullardır.

Assosiasiya qaydalarının analizi. Təhsil məlumatlarının çıxarılması ilə bağlı tədqiqatların əksəriyyəti tez-tez assosiasiya qaydalarının təhlili texnikasından istifadə edilir, çünki o, digər üsullarla müqayisədə daha az təcrübə tələb edir. Buna baxmayaraq, daha sonralar isə tədqiqatçılar tərəfindən analiz üçün klasterləşdirmə və klassifikasiya metodlarından tez-tez istifadə olunduğundan tendensiya dəyişdi. Tez-tez assosiasiya qaydası üçün çoxlu sayda nəticələr əldə edilir, onların əksəriyyəti verilənlərin əldə edilməsində təcrübəsi olmayanlar üçün asanlıqla başa düşülə bilmədiyindən maraqlı deyildir. Düzgün alqoritmləri seçmək üçün tədqiqatçılar ilkin olaraq verilənləri emal etməli və onları tələb olunan nəticəyə uyğunlaşdırmalıdır. Onların tədqiqi təbiətə kiçik miqyasda olduqda, klasterləşdirmə yanaşmasından istifadə edə bilirlər, çünki təsnifat yanaşmasında tələb olunan məlumatların bölünməsi bu texnikada tələb olunmur. Bundan əlavə, tədqiqatçılar həmişə [6]-də istifadə edilən eyni verilənlər bazasından istifadə edərək müxtəlif alqoritmlərlə müqayisə edə bilirlər. Bu, fərqli yanaşmadan istifadə edildikdə eyni nəticələrin əldə edilib-edilməyəcəyini müəyyən etməyə kömək edərdir.

Klassifikasiya. Əsas data mining metodları arasında ən qədim və ən faydalı üsullardan biri klassifikasiyadır. Klassifikasiya çox sadə və ən çox istifadə edilən məlumatların çıxarılması üsuludur. Klassifikasiyanı başa düşmək üçün təlim məlumatlarını bilmək lazımdır. Klassifikasiya prosesinin iki mərhələsi var: təlim üçün modelin hazırlanması; test məlumatlarından istifadə edərək modelin qiymətləndirilməsi.

Klassifikasiya proqnozlaşdırıcı metoddur və alqoritmlər əsasında müxtəlif təsnifat üsulları var:

Statistikaya əsaslanan alqoritmlər: Statistik prosedurlar adətən sadə təsnifatdan daha çox hər bir sinifdə olmaq şansını təmin edən dəqiq fundamental ehtimal modelinə malikdir.

Korrelyasiya Təhlili: Ədədi olaraq ölçülən iki davamlı dəyişənin (məsələn, yaş və çəki) bir-biri ilə əlaqəli olan əlaqə dərəcəsini tapmaq üçün istifadə edilən statistik üsuldur.

Regressiya təhlili: Bu üsul müstəqil dəyişənin asılı dəyişənlə ədədi olaraq necə əlaqələndirilməsini təsvir edir.

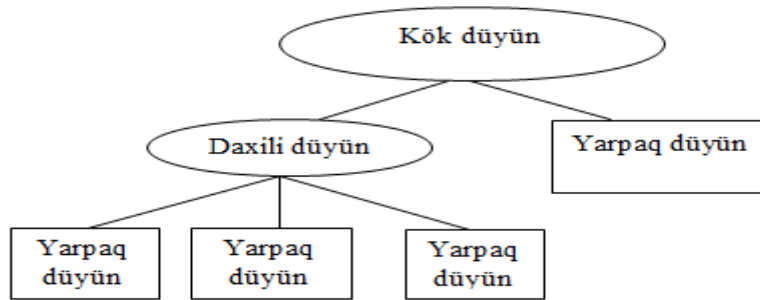
K ən yaxın qonşu: Bu, məsafənin ölçülməsindən asılı olan parametrik olmayan bir üsuldur, bütün mövcud hallar orada saxlanıla bilər və hər dəfə yeni bir vəziyyət daxil olduqda, məsafə funksiyasına əsasən təsnif edilə bilər.

Bayes modeli: Bu üsul tez-tez təkrarlanan texnikadan istifadə edir. Tez-tez təkrarlanma texnikasının mahiyyəti, ehtimalı məlumatlara tətbiq etməkdir. Bayes hesablamaları hipotezin ehtimalı üçün birbaşa gedir. Bu metodun riyazi tənliyi (1)-düsturunda göstərilmişdir.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

X və Y cüt təsadüfi atributlardır. X və Y -in birləşmə ehtimalı $P(X = x, Y = y)$, X dəyişəninin xüsusi x dəyərini, Y dəyişəninin isə y dəyərini qəbul etməsi ehtimalıdır. Şərti ehtimal, başqa bir təsadüfi dəyişənin nəticəsinin məlum olduğu halda təsadüfi dəyişənin xüsusi dəyər alması ehtimalıdır.

Qərar ağacı: Qərar ağacı başqa bir klassifikasiya üsuludur. Qərar ağacı, şəkildə göstəriləyimi kimi, ağac quruluşudur. Ağac kök düyün (oval ilə qeyd olunan birinci düyün), iki və ya daha çox növ ola bilən daxili düyünlərdən (oval ilə işarələnmiş orta düyün) və yarpaq düyünlərindən (düzbucaqlı ilə işarələnmiş) ibarətdir. Bütün daxili qovşaqlarda atributların ifadəsinin dəyərini yoxlayan bölünmələr var. Tədqiqatlardan bəziləri qərar ağaclarının digər klassifikasiya alqoritmləri arasında ən yaxşısı olduğunu, bəziləri isə Bayes Şəbəkəsinin digərlərindən daha yaxşı performans göstərdiyini aşkar edir.



Qərar ağacı

Tələbələrin fəaliyyətinin proqnozlaşdırılması. Proqnozlaşdırmanın məqsədi tələbəni təsvir edən dəyişənin naməlum dəyərini təxmin etməkdir. Təhsildə normal olaraq proqnozlaşdırılan dəyərlər performans, bilik, bal və ya qiymətdir. Bu dəyər kəsilməz (regressiya tapşırığı) və ya diskret dəyər (klassifikasiya tapşırığı) ola bilər. Regressiya təhlili asılı dəyişən ilə bir və ya bir neçə müstəqil dəyişən arasındakı əlaqəni tapır. Klassifikasiya, ayrı-ayrı elementlərə xas olan bir və ya bir neçə xüsusiyyətinə dair kəmiyyət məlumatı əsasında və mövcud etiketlenmiş elementlərin təlim toplusuna əsaslanaraq qruplara yerləşdirildiyi prosedurdur. Tələbə performansını proqnozlaşdırmaq məşin öyrənməsinin vacib istifadəsidir və müxtəlif modellərə (neyron şəbəkələri, Bayes şəbəkələri, qaydalara əsaslanan sistemlər, regressiya və korrelyasiya təhlili) tətbiq edilmişdir. Məşin öyrənmə modeli hər bir tələbə haqqında öyrənmək üçün istifadə olunur ki, bu da ona onların çatışmazlıqlarını müəyyən etməyə və onların təkmilləşdirilməsi yollarını məsələn, daha çox mühazirələrdə iştirak etmək və ya əlavə ədəbiyyatı nəzərdən keçirməklə müəyyən etməyə imkan verir. Bu modellərdən istifadə etməklə klasterləşdirmə, assosiasiya qaydaları və klassifikasiya kimi bir çox bilik növləri aşkar edilə bilər. Aşkar edilmiş biliklərdən tələbələrin müəyyən kursa qəbulu, onlayn imtahanda istifadə olunan ədalətsiz vasitələrin aşkarlanması, tələbələrin nəticə vərəqlərində tələbələrin performansı haqqında proqnoz və s. ilə bağlı proqnoz vermək üçün istifadə oluna bilər.

Elektron təhsildə (qeyri-səlis assosiasiya qaydalarından istifadə etməklə) tələbə fəaliyyətini proqnozlaşdırmaq üçün müxtəlif növ qaydalara əsaslanan sistemlər tətbiq edilmişdir [7]:

– tərtib edilmiş təlim portfelleri əsasında (əsas formativ qiymətləndirmə qaydalarından istifadə etməklə) şagirdin fəaliyyətini proqnozlaşdırmaq;

– tələbənin akademik fəaliyyətinin proqnozlaşdırılması, monitorinqi və qiymətləndirilməsi üçün (qaydaların induksiyasından istifadə etməklə);

– təhsilin veb-əsaslı sistemində daxil edilmiş məlumatlardan çıxarılan xüsusiyyətlər əsasında yekun qiymətləri proqnozlaşdırmaq (assosiasiya qaydalarını tapmaq üçün genetik alqoritmdən istifadə etməklə);

– LMS-lərdə tələbə qiymətlərini proqnozlaşdırmaq (qrammatika ilə idarə olunan genetik proqramlaşdırmadan istifadə etməklə);

– internet əsaslı elektron təhsil sistemlərində (qərar ağacından istifadə etməklə) tələbələrin fəaliyyətini proqnozlaşdırmaq və vaxtında dərslər təqdim etmək;

– tələbələrin onlayn qiymətlərini proqnozlaşdırmaq (ortoqonal axtarışa əsaslanan qayda çıxarma alqoritmindən istifadə etməklə).

Nəhayət, onlayn dərslərdə veb-şagird performansını proqnozlaşdırmaq üçün tələbənin yekun imtahan balını proqnozlaşdırmaq [8] və universitetdə uğur qazanma ehtimallarını proqnozlaşdırmaq lazımdır.

Məşin təlimi. Düzgün karyera istiqaməti tələbələrin performansını və onların motivasiya səviyyəsini artırır. Tələbələrin bacarıqlarını qiymətləndirmək üçün karyera məsləhətinə əsaslanan sistemlər çox vacibdir və buna görə iş götürənlər onlara uyğun işi təyin edirlər [9; s. 9-11]. Bu problemləri həll etmək üçün karyera məsləhəti tələb olunur ki, bu da tələbələrin düzgün kursu

seçməsinə istiqamət verməkdə mühüm rol oynayacaq. Burada, Maşın təlimini (ML – Machine Learning) hesablama intellekt tapşırığını ağıllı şəkildə yerinə yetirmək üçün istifadə ediləcək [10].

ML texnikaları tələbələrin düzgün karyera seçimi ilə bağlı suallarını həll etməyə kömək edir. Təcrübələrdən öyrənmək üçün məlumatların çıxarılması üsullarından və alqoritmlərindən istifadə edilir. ML əsaslı sistem tələbələrin karyera rəhbərliyi ilə bağlı effektiv qərar qəbul etmək üçün Süni İntellekt Texnikaları, Dərin Öyrənmə, Neyron Şəbəkə, Təbii Dil Emalı və s.-dən istifadə edir. Karyera məsləhəti tələbələrə öz maraqlarından və imkanlarından istifadə etmək üçün düzgün karyera yanaşmasını seçmək üçün düzgün qərar qəbul etməyə kömək edir. Üstəlik, təhsil müəssisələrinin və rəqabət mühitinin artması səbəbindən düzgün karyera istiqaməti seçmək tələbələr üçün yorucu işə çevrilir. Zənginləşdirilmiş Maşın Öyrənmə sistemi tələbələrə vəzifədə olan şəxsin imkanlarına uyğun olaraq düzgün kurs kurikulumunu seçməyi tövsiyə edir. ML-yə imkan verən sistemlər tələbələrin imkanlarına uyğun olaraq tələbələrə rəhbərlik etməyə kömək edir. Sistem tələbələrdən aşağıdakı atributlardan ibarət daxiletmə məlumatlarını tələb edir: a) tələbələrin orta ixtisas fənninə olan marağı, b) ixtisas, c) işə qəbul imtahanlarında alınan qiymətlər, d) iş təcrübəsi, e) yerləşdirmə statusu, f) məzuniyyətdən sonra alınan qiymətlər və s. Bu xüsusiyyətlərdən ibarət məlumatlar ibtidai təhsildən, ali təhsildən, məşğulluq idarələrindən və yerli idarəetmə vasitəsi ilə onlayn konvensiyalar, icmalar, sorğular, tələbələrlə üz-üzə qarşılıqlı əlaqə proqramları kimi keyfiyyətli kəmiyyət üsullarından istifadə etməklə toplana bilər.

Dataların təhlili və vizuallaşdırılması. Dataların təhlili və vizuallaşdırılmasının məqsədi faydalı məlumatı vurğulamaq və qərar qəbulunu dəstəkləməkdir. Məsələn, təhsil mühitində, müəllimlər və tələbə rəhbərləri tələbələrin öyrənməsi və dərs fəaliyyəti haqqında ümumi fikir əldə etmək üçün istifadə olunan məlumatları təhlil etməkdə kömək edir. Statistika və vizual məlumat bu vəzifə üçün ən çox istifadə edilən iki əsas texnikadır. Statistika məlumatların toplanması, izahı, təhlili və təqdimatı ilə bağlı riyazi elmdir [11; s. 3]. Təhsil verilənləri ilə istifadə edilən bu təsviri təhlil tələbənin davranışı haqqında xülasələr və hesabatlar kimi qlobal məlumat xüsusiyyətlərini təmin edə bilər. Buna görə də, müəllimlər şərhə asan olan pedaqoji yönümlü statistikaya (ümumi müvəffəqiyyət dərəcəsi, mənimsəmə səviyyələri, tipik yanlış təsəvvürlər, həll olunan tapşırıqların faizi və oxunan material) üstünlük verirlər. Digər tərəfdən, müəllimlər jurnaldakı incə statistik məlumatları yoxlamaq üçün çox çətin və ya şərh etmək üçün çox vaxt aparan hesab edirlər. Təhsil məlumatlarının statistik təhlili (log faylları/verilənlər bazaları) bizə aşağıdakıları deyə bilər: tələbələrin daxil olduğu və çıxdığı yerlər, ən populyar səhifələr, tələbələrin istifadə etdikləri brauzerlər; ziyarətlərin sayı, ziyarətçilərin mənşəyi, hitlərin sayı, müxtəlif dövrlərdə istifadə nümunələri; ziyarətlərin sayı və rüb üzrə müddət, ən yaxşı axtarış şərtləri, elektron təhsil resurslarının yüklənmə sayı; baxılan müxtəlif səhifələrin sayı, müxtəlif səhifələrə baxmaq üçün ümumi vaxt; istifadə xülasələri, həftəlik və aylıq istifadəçi fəaliyyətləri haqqında hesabatlar; sessiya statistikasını və sessiya nümunələri; forumlarda tələbənin qarşılıqlı əlaqəsinə dair statistik göstəricilər; tələbələrin keçə biləcəyi materialın miqdarı, tələbələrin mövzuları öyrənmə ardıcılığı; tələbələrin istifadə etdiyi resurslar, tələbələrin qiymətləndirdiyi resurslar; müzakirə forumlarına töhfələrin ümumi orta göstəriciləri, göndərmələrin miqdarı və cavabların miqdarı, öyrənən-müəllim və ya tələbə-müəllim qarşılıqlı əlaqəsi; tələbənin kursa və ya onun müəyyən hissəsinə ayırdığı vaxt; öyrənənlərin davranışı və vaxt bölgüsü, şəbəkə trafikinin zamanla paylanması; hadisələrin öyrənilmə tezliyi, öyrənmə fəaliyyətinin nümunələri, hadisələrin vaxtı və ardıcılığı, tələbələrin qeydlərinin və xülasələrinin məzmununun təhlili. Statistik təhlil həm də tələbənin bu gün neçə dəqiqə işlədiyini, neçə problemi həll etdiyini və onun düzgün faizini, onun balı və performans səviyyəsi ilə bağlı proqnozu qiymətləndirən hesabatlar əldə etmək üçün çox faydalıdır.

Dataların vizuallaşdırılması insanlara məlumatları anlamaq və təhlil etmək üçün qrafik üsullardan istifadəyə kömək edir. Vizual təsvirlər və qarşılıqlı əlaqə üsulları, istifadəçilərə eyni anda böyük həcmdə məlumatı görmək, araşdırmaq və anlamaq imkanı verir. Müxtəlif təhsil datalarının vizuallaşdırılmasına yönəlmiş bir neçə tədqiqat, məsələn: onlayn forumlarda illik, yarımillik, gündəlik və saatlıq istifadəçi davranışı nümunələri; tam təhsil (qiymətləndirmə) prosesi; riyazi bacarıqları ölçmək üçün verilənlərdə təhlil edilən atributların orta qiymətləri; tapşırıqların tamamlanması, qəbul edilən

suallar, imtahan balı və s. haqqında statistik qrafiklər; tələbələrin sosial, koqnitiv və davranış aspektləri ilə bağlı məlumatların izlənməsi; tələbələrin davamiyyəti, resurslara çıxış, tapşırıqlar və viktorinalar üzrə müzakirələrin və nəticələrin icmalı; tələbələrin və qrupların fəaliyyəti ilə bağlı həftəlik məlumat; tələbənin fərdi anlayışlar və ali təhsil tələbələrinin qiymətləndirmə məlumatları haqqında əsas anlayışında çatışmazlıqlar; tələbənin onlayn öyrənmə mühitləri ilə qarşılıqlı əlaqəsi; tələbələrin onlayn hazırlıq işi, o cümlədən tələbələrin qarşılıqlı əlaqəsi və cavabları, səhvləri, müəllimlərin şərhləri; adaptiv dərslərdə suallar və təkliflər; hərəkət davranışı və şagirdin performansı; ziyarət edilən veb-səhifələrin təhsil yolları, görülən fəaliyyətlər, öyrənmə obyektləri və təhsil yollarının ardıcılığı mövcuddur.

Tələbələrin Modelləşdirilməsi. Tələbə modelləşdirməsinin məqsədi tələbələrin koqnitiv modellərini, o cümlədən onların bacarıqlarının və deklarativ biliklərinin modelləşdirilməsini inkişaf etdirməkdir. Tələbə modellərinin qurulmasını avtomatlaşdırmaq üçün tələbə xüsusiyyətlərini (motivasiya, məmnunluq, öyrənmə üslubları, effektiv status və s.) və öyrənmə davranışını avtomatik nəzərə almaq üçün verilənlərin öyrənilməsindən istifadə edilir. Bayes şəbəkələri tələbə biliyi haqqında proqnozlar vermək, yəni tələbənin müəyyən vaxtda koqnitiv repetitorlar vasitəsilə bir bacarığı bilmə ehtimalı [12]; internet əsaslı təhsil sistemində tələbələrin öyrənmə üslublarını aşkar etmək; tələbənin problemə düzgün cavab verib-verməyəcəyini proqnozlaşdırmaq; İTS-də bacarıqların mənimsənilməsi zamanı tələbənin dəyişən bilik vəziyyətini modelləşdirmək; veb-əsaslı repetitorluq sistemində tələbələrin kömək axtarma davranışından müşahidə olunmayan öyrənmə dəyişənlərini çıxarmaq və özünüintizamın tələbələrin biliyinə və öyrənməsinə təsirini yoxlamaq üçün bilik axtarışı üçün istifadə olunur.

Görüləcək işlər və tədqiqat xətləri. EDM-də nəzərə alınacaq bir çox gələcək iş olsa da, onların arasında ən maraqlı və təsirli olanları qeyd edək. Əslində, bu məqamların bəziləri ilə bağlı bir neçə ilkin tədqiqatlar artıq görünməyə başlayıb.

– EDM alətləri məlumatların əldə edilməsində müəllimlər və ya qeyri-ekspert istifadəçilər üçün daha asan olacaq şəkildə tərtib edilməlidir. Data mining alətləri adətən sadəlikdən daha çox güc və çeviklik üçün nəzərdə tutulub. Mövcud verilənlərin öyrənilməsi alətlərinin əksəriyyəti pedaqoqların istifadə etməsi üçün çox mürəkkəbdir və onların xüsusiyyətləri pedaqoqun etmək istəyəcəyi işin çərçivəsindən kənara çıxır. Məsələn, bir tərəfdən istifadəçilər DM-də mövcud olan geniş metodlar arasından tətbiq etmək istədikləri xüsusi DM metodunu seçməlidirlər. Digər tərəfdən, verilənlərin çıxarılması alqoritmlərinin əksəriyyəti icra edilməzdən əvvəl konfigurasiya edilməlidir. İstifadəçilər yaxşı modellər əldə etmək üçün əvvəlcədən parametrlər üçün müvafiq dəyərlər təqdim etməlidirlər və buna görə də düzgün parametrləri tapmaq üçün istifadəçi müəyyən təcrübəyə malik olmalıdır. Mümkün həll yollarından biri, hər bir tapşırıq üçün standart alqoritmədən istifadə edən istifadəçi alətlərinin və mütəxəssis olmayan istifadəçilər üçün konfigurasiya və icranı sadələşdirmək üçün parametrsiz məlumatların çıxarılması alqoritmlərinin hazırlanmasıdır. EDM alətləri, həmçinin onların nəticələrini müəllimlər və elektron öyrənən dizaynerlər üçün mənalı etmək üçün istifadəsi asan və yaxşı vizuallaşdırma imkanlarına malik daha intuitiv interfeysə malik olmalıdır. EDM istifadəçilərinin hazırda əl ilə etməli olduğu bütün emal funksiyalarını və ya tapşırıqları avtomatlaşdırmaq və asanlaşdırmaq üçün xüsusi ön emal alətlərinin hazırlanması da çox vacibdir.

– Elektron təhsil sistemi ilə inteqrasiya. Data mining aləti daha bir əhəmiyyətli müəllif aləti (kurs yaradıcısı, test yaradıcısı və s.) kimi e-təhsil mühitinə inteqrasiya edilməlidir. Bütün data mining tapşırıqları (ilkin emal, verilənlərin işlənməsi və sonrakı emal) oxşar interfeysə malik bir proqramda həyata keçirilməlidir. Bu yolla, EDM alətləri müəllimlər tərəfindən daha geniş istifadə ediləcək və məlumatların öyrənilməsi üsulları ilə əldə edilən rəy və nəticələr təkrarlanan qiymətləndirmə prosesindən istifadə etməklə asanlıqla və birbaşa e-tədris mühitinə tətbiq edilə bilər.

– Məlumatların və modellərin standartlaşdırılması. Müəyyən bir kursa aid maining məlumatları üçün cari alətlər yalnız onların tərtibatçıları üçün faydalı ola bilər. Hər hansı bir təhsil sistemində tətbiq oluna bilən ümumi alətlər və ya təkrar istifadə vasitələri yoxdur. Beləliklə, ilkin emal, axtarış və sonrakı emal tapşırıqları ilə yanaşı, giriş məlumatlarının və çıxış modelinin standartlaşdırılması

lazımdır. Bəzi müəlliflər [13] verilənlərin spesifikasiyası kimi XML-dən istifadə etməyi təklif etmişlər. Digər müəlliflər [14] statistik və data mining modelləri üçün aparıcı standart olan PMML-dən (Predictive Modeling Markup Language) istifadə etmişlər. Lakin OWL (Ontology Web Language) və RDF (Resource Description Framework) kimi ontoloji spesifikasiya dillərindən və SCORM (Shareable Content Object Reference Model) kimi e-təhsil üçün standart metadata istifadə edərək domen biliklərini və semantikasını birləşdirmək də lazımdır. Bu xətdə, hal-hazırda, çoxlu təhsil data bazalarını təmin edən və təhlili asanlaşdıran yalnız bir ictimai təhsil data bazası olan PSLC DataShop var. Bununla belə, bütün bu məlumatlar İntellektual Repetitorluq Sistemlərindən əldə edilir, ona görə də digər təhsil mühitlərindən daha çox ictimai verilənlər bazasına sahib olmaq lazımdır. Bu yolla, müxtəlif məlumatların çıxarılması alqoritmlərini qiymətləndirmək üçün xüsusi təhsil etalon verilənlər bazasından istifadə oluna bilər.

– Ənənəvi mining alqoritmləri təhsil kontekstini nəzərə almaq üçün tənzimlənə bilər. Data mining üsulları təhsil mininglərinə tətbiq edildikdə semantik məlumatlardan istifadə etməlidir. Bu, təhsil sahəsinə dair bilikləri data mining alqoritmlərinə inteqrasiya edən daha effektiv mining alqoritmlərinə ehtiyac olduğunu göstərir. Məsələn, bəzi müəlliflər [15] Öyrənmə İdarəetmə sistemi tərəfindən qeydə alınan trekin semantikasını təsvir etmək və onları proqnozlaşdırıcı ssenaridə müəyyən edilmiş müşahidə ehtiyacı ilə əlaqələndirmək üçün xüsusi istifadə izləmə dilini (UTL) təklif etmişlər. Təhsilə xas mining üsulları tədris dizaynını və pedaqoji qərarları əhəmiyyətli dərəcədə təkmilləşdirə bilər və semantik internetin məqsədi təhsil mühitlərində məlumatların idarə edilməsini asanlaşdırmaq.

Nəticə. Data mining təhsil müəssisələrində mühüm əhəmiyyət kəsb edir. Data mining üsullarının istifadəsi ilə əldə edilən biliklər, tələbənin təhsildəki performansını yaxşılaşdıracaq və irəliləyəcək, uğurlu və effektiv qərarlar qəbul etmək üçün istifadə edilə bilər. Bu məqalə EDM ilə bağlı ən müasir vəziyyətin nəzərdən keçirilməsidir və bu sahədə bu günə qədər ən uyğun işi araşdırır. Hər bir tədqiqat yalnız istifadə olunan məlumatların növü və DM üsulları ilə deyil, həm də həll etdikləri təhsil tapşırığının növü ilə təsnif edilmişdir. EDM, elektron təhsil, adaptiv hipermedia, ağıllı repetitorluq sistemləri, veb-məlumatlar, məlumatların öyrənilməsi və s. daxil olmaqla, bir neçə yaxşı qurulmuş tədqiqat sahələri ilə bağlı gələcək tədqiqat sahəsi kimi təqdim edilmişdir. Biz EDM-in Beynəlxalq Konfranslarda və Jurnallarda hər il nəşr olunan məqalələrin sayının artmasında necə sürətlə inkişaf etdiyi görürük. Beləliklə, demək olar ki, EDM indi yeniyetməlik dövrünə yaxınlaşır, yəni artıq ilk dövrlərində deyil, hələ yetkin bir sahə də deyil. Əslində, biz bəzi maraqlı gələcək xətləri təsvir etdik, lakin bunun daha yetkin bir sahəyə çevrilməsi üçün tədqiqatçıların indiki çoxsaylı fərdi təkliflər və xətlər əvəzinə daha vahid və birgə tədqiqatlar hazırlaması da lazımdır. Beləliklə, məlumatların öyrənilməsinin təhsil mühitinə tam inteqrasiyası reallığa çevriləcək və tam operativ tətbiqetmələr (həm kommersiya, həm də pulsuz) təkcə tədqiqatçılar və tərtibatçılar üçün deyil, həm də xarici istifadəçilər üçün əlçatan olacaq.

ƏDƏBİYYAT

1. Romero C., Ventura S., Pechenizkiy M., and Baker R. Handbook of educational data mining. CRC Press, 2010, p. 503.
2. Goyal M. and Vohra R. Applications of Data Mining in Higher Education. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012, p. 113-120.
3. Huebner R. A survey of educational data mining research. Research in Higher Education Journal, 2012, p. 1-13.
4. Mythili M.S., Mohamed Shanavas A.R. An Analysis of students' performance using classification algorithms. IOSR, Journal of Computer Engineering, Volume 16, Issue 1, January 2014, p. 63-69.
5. Lakshmi Prabha S., Mohamed Shanavas A.R. Educational data mining applications. Operations Research and Applications: An International Journal (ORAJ), Vol. 1, No. 1, August 2014, p. 23-29.
6. Pechenizkiy M., Calders T., Vasilyeva E., De Bra P. Mining the student assessment data. lessons drawn from a small scale case study. In: Educational Data Mining 2008 (2008), p. 187-191.
7. Nebot A., Castro F., Vellido A., Mugica F. Identification of fuzzy models to predict students performance in an e-learning environment. In International Conference on Web-based Education, Puerto Vallarta, 2006, p. 74-79.

8. Pritchard D., Warnakulasooriya R. Data from a Web-based Homework Tutor can predict Student's Final Exam Score. In World Conference on Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, 2005, p. 2523-2529.
9. Viheräköski J. Strengths in career development: Modeling strength-based career counseling through reflecting customer experience. 2020, p. 105.
10. Yang C., Huan S., & Yang Y. A practical teaching mode for colleges supported by artificial intelligence. *International Journal of Emerging Technologies in Learning (IJET)*, 2020, 15(17), p. 195–206.
11. Freedman D., Purves R., Pisani R. Statistics, 4 th Edition. W.W. Norton & Co. 2007, p. 69.
12. Baker R., Corbett A.T., Aleven V. Improving contextual models of guessing and slipping with a truncated training set. In International Conference on Educational Data Mining, Montreal, Canada, 2008, p. 67-76.
13. Shen R., Han P., Yang F., Yang Q., Huang J. Data mining and case-based reasoning for distance learning. *Journal of Distance Education Technologies*, 2003, p. 1,3, 46–58.
14. Ventura S., Romero C., Hervas C. Analyzing rule evaluation measures with educational datasets: a framework to help the teacher. In International Conference on Educational Data Mining, Montreal, Canada, 2008, p. 177-181.
15. Iksal S., Choquet C. Usage Analysis Driven by Models in a Pedagogical Context. Workshop on Usage Analysis in Learning Systems, In 12th International Conference on Artificial Intelligence, Amsterdam, 2005, p. 1-8.

APPLICATION OF DATA MINING METHODS IN EDUCATION

F.T.Agayev, N.I.Bahadurzada

*Ministry of Science and Education Republic of Azerbaijan
Institute of Information Technologies*

Abstract. Education Data Mining is one of the emerging fields of modern times, it develops methods to explore different types of data obtained by students from the field of education, and these methods help to understand students and their learning environment. Data mining plays an important role in education, especially when evaluating student performance in an online learning environment. Education Data Mining uses computational approaches to analyze educational data. The purpose of this article is to introduce the different user groups, types of educational environments and the information they provide in Education Data Mining. Then, some of the most promising directions of future research and tasks solved by data mining methods in the educational environment are discussed.

Keywords. *Educational Data Mining, Student Performance Prediction, Data Mining.*

ПРИМЕНЕНИЕ МЕТОДОВ ДАННЫХ В ОБРАЗОВАНИИ

Ф.Т.Агаев, Н.И.Бахадурзада

*Министерство Науки и Образования Азербайджанской Республики
Институт информационных технологий*

Резюме. Интеллектуальный анализ данных в образовании — одна из новых областей современности, она разрабатывает методы для изучения различных типов данных, полученных учащимися в сфере образования, и эти методы помогают понять учащихся и их учебную среду. Интеллектуальный анализ данных играет важную роль в образовании, особенно при оценке успеваемости учащихся в онлайн-среде обучения. Education Data Mining использует вычислительные подходы для анализа образовательных данных. Цель этой статьи — представить различные группы пользователей, типы образовательных сред и информацию, которую они предоставляют при интеллектуальном анализе данных для образования. Затем обсуждаются некоторые наиболее перспективные направления будущих исследований и задачи, решаемые методами интеллектуального анализа данных в образовательной среде.

Ключевые слова: *Образовательный интеллектуальный анализ данных, прогноз успеваемости учащихся, интеллектуальный анализ данных.*

Daxil olub: 30.10.2022