SATBAYEV UNIVERSITY

Computing & Engineering

Volume 2 (2024), Issue 2, 35-41

https://doi.org/10.51301/ce.2024.i2.06

Intellectual scoring and credit risk

Ye. Ybrayev, Zh. Alibiyeva*

Satbayev University, Almaty, Kazakhstan

**Corresponding author: zh.alibiyeva@satbayev.university*

Abstract. This article explores innovative approaches to improving credit scoring systems and managing credit risks in the banking sector. The proposed system utilizes machine learning methods to perform a comprehensive analysis of clients' financial history, market conditions, and macroeconomic indicators, thereby enabling a more accurate assessment of credit risk. The study includes a comparative analysis of various models, demonstrating their practical applicability and efficiency in a rapidly changing financial environment. Experimental results indicate that the application of the proposed algorithms significantly enhances forecasting accuracy and reduces financial losses. The research highlights the potential for implementing adaptive analytical tools that support prompt and well-founded managerial decision-making in banks.

Keywords: *artifical intelligence, machine learning, logistic regression, dataset, database, integration*.

1. Кіріспе

Қазіргі банктік жүйе клиенттердің төлем қабілеттілігін дәл талдау қажеттілігіне тап болып отыр, бұл заманауи ақпараттық технологиялар мен мәліметтерді талдау әдістерін қолдануды талап етеді. Машиналық оқыту технологиясының дамуы мен қолжетімді мәліметтердің көлемінің артуымен дәстүрлі несиелік скоринг әдістері орнына, тұтынушылардың төлем қабілеттілігіне әсер ететін көптеген факторларды ескере алатын икемді және дәл модельдер басымдыққа ие болуда [1].

Осы жұмысқа машиналық оқыту әдістеріне негізделген интеллектуалды несиелік скоринг жүйесін құру ұсынылады. Жүйе клиенттердің қаржылық тарихы, нарықтық жағдайлар мен макроэкономикалық көрсеткіштерді кешенді түрде талдай отырып, несиелік тәуекелді дәлірек бағалауға мүмкіндік береді. Мұндай тәсіл тек қана болжамдардың дәлдігін арттырып қоймай, сонымен қатар, әлеуетті қиындық туғызуы мүмкін клиенттерді уақытында анықтап, қаржылық шығындарды азайтуға септігін тигізеді.

Несиелік скоринг мәселесін шешуде қолданылатын әртүрлі машиналық оқыту алгоритмдерінің салыстырмалы талдауына ерекше көңіл бөлінеді [2]. Қаржы нарығының жылдам өзгерістеріне байланысты, өзгерістерге тез бейімделе алатын және жаңа нарықтық тенденцияларға шұғыл әрекет ете алатын жүйенің маңызы зор. Сондықтан жұмыста нақты банктік процестерде қолдануға ыңғайлы және жоғары адаптивті модельдер қарастырылады.

Несиелік скоринг саласындағы заманауи зерттеулер бірнеше машиналық оқыту алгоритмдерін біріктіретін гибридті модельдердің болжамдық дәлдікті арттыратынын көрсетеді. Осы жұмыста тұтынушылардың төлем қабілеттілігіне әсер ететін түрлі факторлар арасындағы күрделі байланысты ескеретін оңтайлы шешімдерді анықтау мақсатында, осындай модельдердің егжей-тегжейлі талдауы жүргізіледі. Сонымен қатар, экономикалық тұрақсыздық жағдайында модельді жаңа нарықтық жағдайларға бейімдеу механизмдерін әзірлеу мәселесі де өзектілігін сақтайды.

Техникалық аспектілерден бөлек, жұмыс ұсынылған жүйені банктік ортада жүзеге асыру мәселелеріне де назар аударады. Атап айтқанда, жүйенің банктердің бар ИТ-инфрақұрылымдарымен интеграциясы және мәліметтердің қауіпсіздігі мен құпиялылығын қамтамасыз ету мәселелері талқыланады. Қорытынды пайдаланушылар үшін интуитивті түсінікті интерфейс құрудың маңыздылығы айқын көрсетіліп, бұл аналитиктер мен менеджерлерге модель нәтижелерін тиімді интерпретациялап, негізделген шешімдер қабылдауға мүмкіндік береді [3].

Сонымен қатар, жүргізілген зерттеу әрі қарай дамып, кеңейтілу әлеуетіне ие. Ұсынылған жүйені іске асыру несиелік скоринг аясында шешім қабылдау үдерісін жақсартумен шектелмей, басқа қаржылық көрсеткіштерді болжауға бағытталған қосымша аналитикалық модульдерді енгізудің негізі бола алады.

2. Зерттеу әдістері мен материалдары

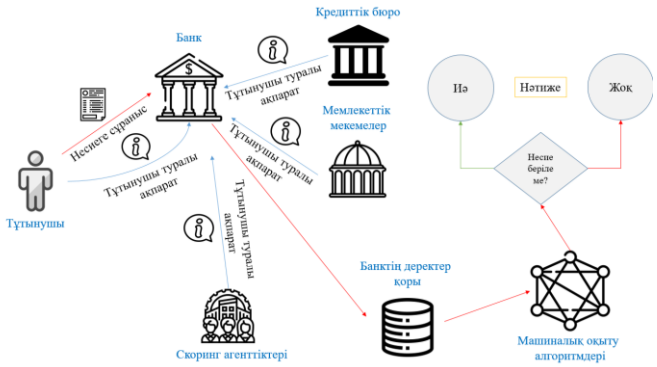
2.1. Несие скорингінің жұмыс жасау принципі және сұлбасы

Әр тұтынушының банк алдындағы талабы оның банк алдындағы борышы мен берешектеріне байланысты индивидуалды болып келеді. Сол себептен де скорингтің жұмыс жасауы әркімге әрқалай нәтижесін шығарады.

Қазіргі таңда, банкке тіркелгенде, не болмаса сол банктің өнімін, банк картасын аштырғанда барлығымыздан келісім сұралады, ол келісім сіздің сол банкке барлық деректеріңізді өңдеу мен сақтауға құқылы екенін растайтын келісімшарт болып табылады. Осы жерде туындайтын үлкен мәселе мынандай сұрақтың

туындауына алып келеді: «Банк бұл деректерді не мақсатта сақтайды?».

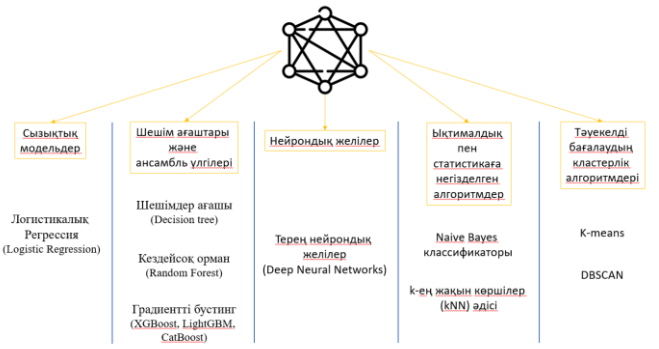
Егер логикалық тұрғыда қарайтын болсақ, адамдар бірін бірі танымай араларында ақша алмасу болуы екі талай емес пе? Дәл осы принципте, банк сіз туралы деректерді жинай арқылы сізді тануға кіріседі, ал ол деректер скоринг кезінде қолданылады, сол деректерді өңдей арқылы сізге несие беріле ме, жоқ па, оны банк жүйесінде бағдарламаланған машиналық оқыту алгоритмдері шешім шығарады (1-сурет).



Сурет 1. Несие скоринг жүйесінің жалпылама сұлбасы

2.2. Несиелік скоринг моделын бағдарламалау алгоритмдері

Скорингтің бұл түрін бағдарламалау барысында қолданылатын алгоритмдер бір емес, алайда, осы алгоритмдердің дұрыс жұмыс жасауын біз ол модельді бағалай арқылы ғана білеміз. Негізі, программалық кодтар ғаламторда ашығынан жарияланған, және осы кодтарды сараптамадан өткізе мына алгоритмдер кеңінен қолданылатынын айта аламыз (2-сурет).



Сурет 2. Несие скорингін бағдарламалауда кеңінен қолданылатын алгоритмдер

Маңызды айта кететін ақпарат ретінде жергілікті Қазақстандағы банктердің қолданатын алгоритмдерін айта кетуге болады. Мысалы Kaspi Bank және Halyk Bank мекемелерінде шешімдер ағашы үлгісі кеңінен қоладынлуда. CatBoost, XGBoost секілді градиентті бустинг әдістері арқылы тұтынушыға несие беруге болады ма, жоқ па шешім шығаруда [4].

2021-жылдан бастап Банк ЦентрКредит банкі де несиелік скорингті машиналық оқыту негізінде іске асырып несие портфелін 20-30%-ға дейін арттырып отыр. Банк ЦентрКредит модельдерді әзірлеу, валидациялау және орналастыру процестерін автоматтандыратын өзінің машиналық оқыту платформасын әзірледі, бұл машиналық оқыту үлгілерінің өмірлік циклін тиімді басқаруға мүмкіндік береді. Kaspi Bank пен Halyk Bank секілді бұл банкте де градиентті бустинг алгоритмдері (XGBoost) кеңінен қолданыста. Бұл әдістер деректердің үлкен көлемін өңдеуге және күрделі заңдылықтарды анықтауға мүмкіндік береді, бұл несиелік тәуекелдерді дәлірек бағалауға ықпал етеді [5].

Jusan Bank несиелік скорингті қоса алғанда, бизнес-процестерін оңтайландыру үшін жасанды интеллект және машиналық оқыту технологияларын белсенді түрде енгізуде. Банк өзі қолданатын нақты алгоритмдерді ашып көрсетпесе де, қызмет көрсету сапасын және клиенттермен және серіктестермен өзара әрекеттесу тиімділігін арттыруға мүмкіндік беретін машиналық оқыту саласында өзінің инновациялық шешімдерін әзірлейтіні белгілі.

Нәтиже ретінде осы банктің келесі жетістектерін айта кетуге болады:

2022 жылы банк өзінің клиенттер қорын 52%-ға дейін арттырып жалпылай келгенде 3 млн. тұтынушымен өзінің өнімімен бөлісіп отыр.

Банк проблемалық несиелер үлесін айтарлықтай төмендетіп, 2021 жылдың басындағы банктер рейтингіндегі тоғызыншы орыннан сол жылдың соңына қарай төртінші орынға көтерілді.

2.3. Несиелік тәуекелді бағалау үшін машиналық оқыту алгоритмдерінің салыстырмалы талдауы

Жоғарыда айта кеткеніміздей, дәл осы жүйені бағдарламалауда бір емес, бірнеше алгоритмдерді қолдана аламыз, тіпті оларды комбинациялап нәтиже сапасын жақсартуға болады. Әрине, іске асыру барысында, комбинациялау қосымша ресурс, уақыт алуы мүмкін, алайда, мүмкіншілік болса, нәтижені одан сайын жақсарту банктің тәуекелін айтарлықтай азайтады.

Жобалау барысында, проект бірнеше алгоритм бойынша іске асты, олардың бірі – логистикалық регрессия.

Логистикалық регрессия – бақыланатын машиналық оқыту әдісі және берілген пән үшін оқиғаның орын алу ықтималдығын бағалау үшін пайдаланылуы мүмкін көп регрессия түрі (ауру/сау, несиені өтеуге қабілетті/қабілеттілігі төмен және т.б.). Жіктеу мәселесінде логистикалық регрессия «бинарлы» деп аталады, ол тәуелді айнымалы екілік болғанда қолданылады, яғни ол тек екі мәнді қабылдай алады, мысалы, иә немесе жоқ, шын немесе жалған, 1 немесе 0 және т.б. Логистикалық регрессияның мәні мынада: тәуелсіз айнымалылар мәндерінің кеңістігі (n-өлшемді кеңістік) қандай да бір сызықтық шекара арқылы сыныптарға сәйкес келетін екі аймаққа бөлінуі мүмкін [6].

Қандай да бір сызықтық шекара деп гипержазықтық деп аталатын тәуелсіз айнымалылардың сызықтық комбинациясын түсінеміз. Екі өлшемді жағдайда бұл қисықсыз түзу сызық. Үш өлшемді жағдайда бұл ұшақ және т.б. Нүкте гипержазықтықтан неғұрлым алыс болса, оның пәндік облыстың сәйкес класына жататындығы соншалықты жоғары болады [7].

Гипержазықтықты сипаттайтын теңдеу:

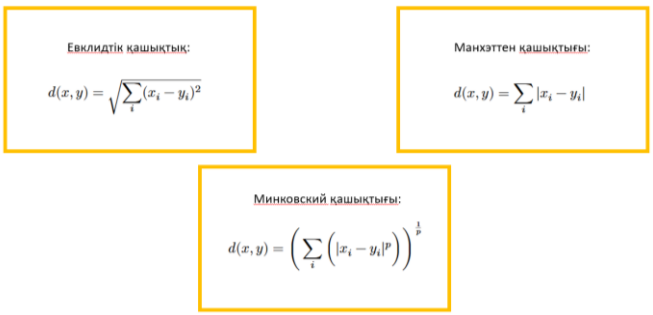
$$w^T x_i + b = 0$$

$w_1x_1 + w_2x_2 + \ldots w_ix_i + b = 0$

$\sum_{i=1}^{n} w_i x_i + b = 0 (w \in R, b \in R)$

мұндағы w_i - нүкте арқылы өтетін сызықтардың бүкіл тобын анықтайтын коэффициенттер w қатынасы сызықтың осьтерге бұрышын анықтайды b нөлден тыс коэффициент түзудің нөлден өтпеуіне мүмкіндік береді. Басқаша айтқанда осьтерге бейімділік өзгермейді, яғни b параллель түзулер тобын анықтайды, n белгісіз айнымалылар саны. Вектордың геометриялық мағынасы (w_i, w_n) гипержазықтық үшін қалыпты болып келеді.

Логистикалық регрессиядан бөлек, бағдарламалау мен модельдеуге k-Nearest Neighbours (KNN) алгоритмін қоладнуға да болады. k - Nearest Neighbours (KNN) – классификация және регрессия есептерін шешу үшін қолданылатын машиналық оқыту алгоритмінің түрі. KNN қарапайым идеяға негізделген: бір сыныптың объектілері әдетте мүмкіндіктер кеңістігінде бір-біріне жақынырақ болады [8]. KNN принципіне негізделген, мұнда k «ең жақын көршілер» санын білдіреді. Жаңа бақыланатын тип (немесе нысан) жіктеуді қажет еткенде, KNN алгоритмі оқу деректер жинағынан осы нысанның k жақын көршілеріне қарайды. Содан кейін ол нысанға сол көршілер арасында жиі кездесетін сыныпты тағайындайды. Егер k = 1 болса, жаңа нысан ең жақын көршісінің класына тағайындалады. Регрессия контекстінде KNN жаңа мүмкіндікке k жақын көршілерінің орташа мәнін тағайындайды. Бірақ біз KNN-ді жіктеуге шешім қабылдағандықтан, біз регрессияға тоқталмаймыз. Жақын көршілерді анықтау үшін әртүрлі қашықтық өлшемдерін қолдануға болады (3-сурет).



Сурет 3. Қашықтық өлшеу формулалары

мұндағы x және y екі экземпляр, x_i және y_i сәйкесінше x және y үшін мүмкіндіктің i мәндері, ал p жақын көршілердің дәрежесін анықтайтын параметр болып табылады [9].

Жаттығу жиынындағы барлық көршілерге дейінгі қашықтықтарды есептегеннен кейін KNN k арақашықтықтары минималды болатын көршілерді таңдайды. k саны мен қашықтық метрикасын ескере отырып, KNN жаңа бақылауларды ең жақын көршілері k арасында жиі кездесетін класстардың негізінде жіктейді.

Векторлық қолдау машинасы (SVM). Векторлық қолдау машинасы (SVM) негізінен жіктеу мәселелерін шешу үшін пайдаланылатын машиналық оқыту алгоритмінің бір түрі болып табылады, бірақ оны регрессия мен шектен шығуды анықтауға да қолдануға болады. SVM мақсаты - деректердің екі класын жақсы ажырататын жоғары өлшемді кеңістікте гипержазықтықты табу.

Гипержазықтық - бұл класстар арасындағы алшақтықты құра отырып, енгізу мүмкіндіктерінің кеңістігін бөлетін сызық. SVM деректер нүктелерін олардың белгілі бір сыныптағы мүшелігіне негізделген тиімді түрде бөлетін оңтайлы гипержазықтықты табуға бағытталған – бұл класс «0» немесе класс «1» болып бөлінуі мүмкін. Егер де біз осы класстарды деректерімізді бөлетін екі өлшемді кеңістік контекстіндегі сызық ретінде елестетсек. Мысалы:

$B_0 + B_1X_1 + B_2X_2 = 0$

мұндағы сызықтың көлбеуін және кесу нүктесін B_0 анықтайтын B_1 және B_2 коэффициенттері оқыту алгоритмі арқылы табылады, ал X_1 және X_2 екі кіріс айнымалысы болып табылады [10].

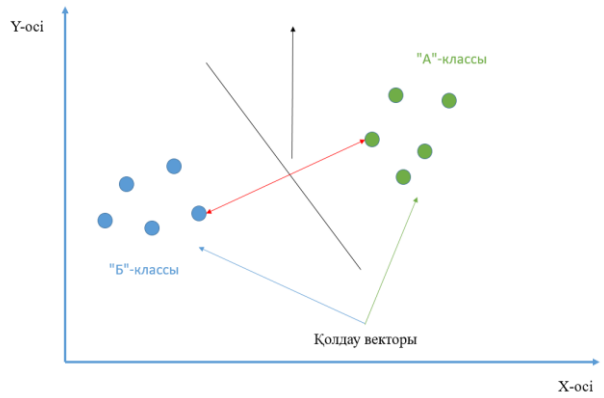
Гипержазықтық сызық жіктеуге мүмкіндік береді. Сызықтың теңдеуін пайдалана отырып, біз кіріс мәндерін қосып, жаңа нүктенің сызыққа қатысты орнын анықтай аламыз. Ол үшін бізде 4 нұсқа болады:

Егер жаңа нүкте түзуден жоғары болса, теңдеу нөлден үлкен мән береді және бұл нүктені бірінші класқа (А класына) жатқызуға болады;

Егер жаңа нүкте түзуден төмен болса, теңдеу нөлден кіші мән береді және нүктені екінші класқа (В класына) жатқызамыз;

Жаңа нүкте түзуге жақын болса, теңдеу нөлге жақын мән береді және нүктені классификациялау қиындай түседі;

Егер мән үлкен болса, бұл модель өзінің болжамына сенімдірек екенін көрсетеді.



Сурет 4. Векторлық қолдау машинасы мысалы

Сызық пен ең жақын деректер нүктелерінің арасындағы қашықтық «қор» ретінде анықталады. Екі классты тиімді ажырата алатын идеалды немесе оңтайлы сызық ең үлкен қорға ие. Мұндай сызық «максималды жиегі бар гиперплан» деп аталады. Қор түзуден ең жақын нүктелерге дейінгі перпендикуляр қашықтық ретінде анықталады. Сызықты анықтау және классификаторды құру кезінде тек осы нүктелер маңызды. Мұндай нүктелер «тірек векторлары» деп аталады. Олар гипержазықтықты қолдайды немесе анықтайды. Гипержазықтық оқу деректерінен қорды барынша арттыратын оңтайландыру әдісі арқылы таңдалады. Тәжірибеде деректер көбінесе хаотикалық болып табылады және оны гипержазықтықпен

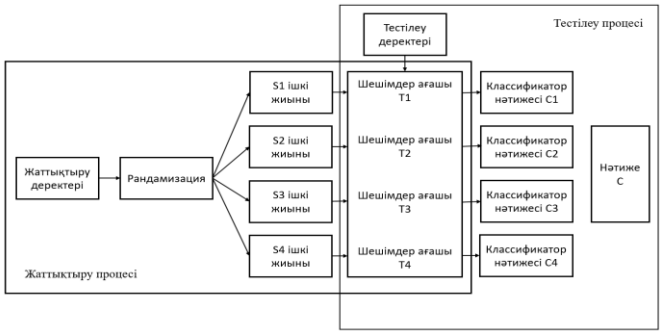
толығымен бөлуге болмайды. Осыған байланысты, біз класстарды бөлетін сызық түріндегі қорды барынша арттыруға шектеуді аздап босаңсытамыз. Мұндай классификатор әдетте «жұмсақ байланысқан классификатор» деп аталады. Бұл өзгерту жаттығу деректер жинағындағы кейбір нүктелерге бөлу сызығының шекарасын бұзуға мүмкіндік береді [11].

Кездейсоқ орман (Random Forest). Кездейсоқ орман - дәлірек және тұрақты болжам жасау үшін бірнеше шешім ағаштарын біріктіретін ансамбльді оқыту әдісі. Әдістің мәні әртүрлі шешім ағаштарының үлкен санын құрастыру, содан кейін олардың болжамдарын орташалау болып табылады. Әрбір ағаш басқалардан тәуелсіз салынған [12]. Әрбір ағашты құру үшін келесі кездейсоқ элементтер қолданылады:

Жүктеу: әрбір ағашты құру үшін кері ізі бар бастапқы деректер үлгісінен ішкі үлгі таңдалады. Бұл кейбір типтер ішкі үлгіде бірнеше рет пайда болуы мүмкін, ал басқалары мүлдем көрінбеуі мүмкін дегенді білдіреді.

Кездейсоқ белгілердің ішкі кеңістігі. Шешім ағашының әрбір түйінінде мүмкіндіктер үлгісі бөлу үшін оңтайлы мүмкіндікті таңдаған кезде мүмкіндіктердің кездейсоқ жиынымен шектеледі.

Дегенмен, Random Forest - пен жұмыс істегенде әрбір ағашты бөлек өңдеудің қажеті жоқ екенін атап өткен жөн. Біз модельді деректерге үйретеміз және ол барлық ағаштарды автоматты түрде жасайды және біріктіреді.



Сурет 5. Кездейсоқ орман классификаторының концептуалды негізі

Шешім ағашындағы бөлудің «сапасын» өлшеу үшін пайдалануға болатын бірнеше көрсеткіштер бар және бұл көрсеткіштер кездейсоқ орманда да қолданылады. Ең көп таралған екі өлшем - энтропия және Джини коэффициенті. Екі классификация да бинарлы болып табылады және түйіннің «тазалық» дәрежесін өлшейді [13].

1) Энтропия (екілік классификация жағдайында):

$$E(p) = -p * \log2(p) – (1 - p) * \log2(1 - p)$$

2) Джини коэффициенті:

$$G(p) = 1 – (p^2 + (1 - p)^2)$$

мұндағы *P* – түйіннен кездейсоқ таңдалған элементтің 1 - классқа жататын болу ықтималдығы. Осылайша, егер түйіндегі барлық элементтер бір класқа жататын болса (яғни түйін «таза» болса), энтропия мен Джини коэффициенті 0-ге тең болады. Егер деректер екі класс арасында біркелкі үлестірілсе, олар максималды мәнге жетеді (егер бинарлы классификация болса онда 0.5-ке тең болады). Кездейсоқ орман контекстінде әр кезеңде ең тиімді бөлуді анықтау үшін энтропия және Джини коэффициенті қолданылады. Бұл бөлім еншілес түйіндердегі энтропиялардың өлшенген сомасын немесе Джини коэффициентін азайту үшін таңдалған. Бұл деректерді жақсы тарататын шешім ағаштарын жасауға көмектеседі және жалпы ансамбль үшін дәлірек болжам жасауға әкеледі. Несиелік скорингтегі кездейсоқ орман мүмкіндіктердің үлкен санын өңдеуде артықшылықтарды, жетіспейтін деректер мен мүмкіндіктер масштабына сенімділікті және теңгерімсіз деректермен жұмыс істеу кезінде тиімділікті ұсынады. Бұл алгоритм несиелік шешім қабылдау кезінде негізгі факторларды бөліп көрсетеді және артық төлем жасау қаупін азайтады. Дегенмен, ол нәтижелерді интерпретациялауда қиындықтар туғызады, есептеу қарқындылығымен ерекшеленеді және болжамның тұрақсыздығы мен сирек оқиғаларды болжау қиындығын көрсете алады. Сондай-ақ, шулы функциялар болған кезде тағы оқыту қаупі бар [14].

3. Зерттеу нәтижелері және оларды талқылау

Мақалада машиналық оқыту үлгілері, соның ішінде k - nearest neighbours (KNN), кездейсоқ орман, векторлық қолдау машинасы (SVM) және логистикалық регрессия қолданылды және әртүрлі көрсеткіштер талданды. Деректер алдын ала өңделді, соның ішінде масштабтау, категориялық мүмкіндіктерді кодтау және шектен тыс мәндерді жою. Содан кейін деректер үлгілерді сенімді бағалауды қамтамасыз ететін қатынаста оқу жинағына және сынақ жиынына бөлінді. Әр модель сәйкес гиперпараметрлерді пайдалана отырып, оқу жинағында оқытылды. Содан кейін сынақ үлгісі үшін болжамдар алынды және мақсатты айнымалының шынайы мәндерімен салыстырылды. Эксперимент нәтижелері барлық қарастырылған модельдер несиелік скоринг мәселесін шешуде тиімділіктің сол немесе басқа дәрежесін көрсететінін көрсетті. Атап айтқанда, KNN және RandomForest үлгілері жоғары дәлдік пен F1-өлшем мәндеріне қол жеткізді, бұл олардың төмен және жоғары несиелік тәуекелді тұтынушыларды дұрыс жіктеу қабілетін көрсетеді. Екінші жағынан, SVM және логистикалық регрессия үлгілері де қолайлы нәтиже көрсетті, бірақ олардың өнімділігі алдыңғы үлгілерге қарағанда біршама төмен болды. Жалпы алғанда, алынған нәтижелерге сүйене отырып, KNN және Random Forest модельдері кредиттік скорингтік тапсырма үшін ең қолайлы алгоритмдер болып табылады деген қорытынды жасауға болады. Дегенмен, бұл нәтиже барлық деректер жиыны үшін орындалмауы мүмкін [15].

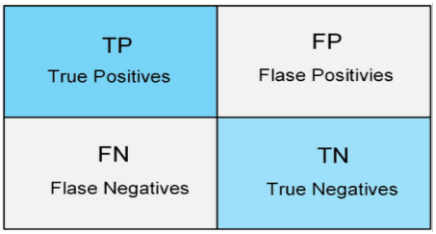
Несиелік скоринг жұмысының ерекшелігі – нашар тұтынушыны жақсы тұтынушы етіп көрстеу банкке шығын көбірек алып келеді, ал керісінше жағдайда шығын болуы мүмкін, алайда болу қаупі біразға төмер болып келеді. Бұл банк тәуекелділігін өте қатты азайтады, сол үшін де, дәлдіккі арттыру мақсатында мынандай ұғымдар пайда болады (6-сурет).

1) TP (True Positive) – шынымен позитивті;

2) FP (False Positive) – жалған оң. I типті қате;

3) FN (False Negative) – жалған теріс нәтиже. II типті қате;

4) TN (True Negative) – шынайы теріс.



Сурет 6. Шатасу матрицасы (Confusion Matrix)

Accuracy – дұрыс жіктелген несиелердің үлесі. Ең қарапайым метрика, бірақ үлгінің жалғыз метрикасы болмауы керек, әсіресе әртүрлі класстардың өкілдері әртүрлі ықтималдықтармен (балансталмаған таңдау) орын алған кезде. Дәлдік келесі формула бойынша есептеледі [16]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision – классификатор оң деп атайтын, шын мәнінде оң болатын объектілердің үлесі. Келесі формула арқылы есептеледі [17]:

$$\Pr ecision = \frac{TP}{TP + FP}$$

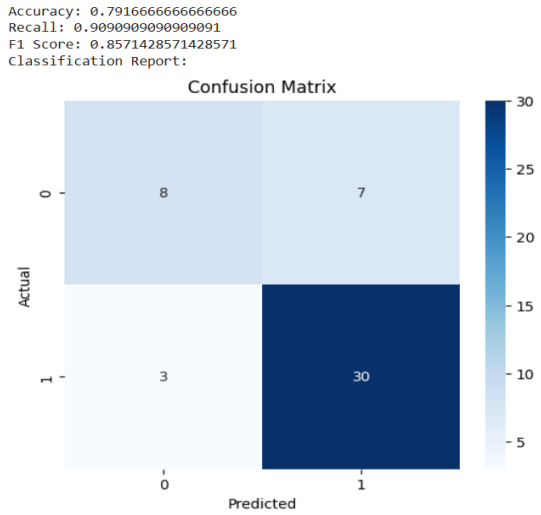
Recall - алгоритм арқылы табылған барлық оң сынып объектілерінен оң сынып объектілерінің үлесі. Ол келесі формула арқылы есептеледі [18]:

$$\Re call = \frac{TP}{TP + FN}$$

*F*1 бағалау – Precision мен Recall мәндерінің гармоникалық ортасы [19]:

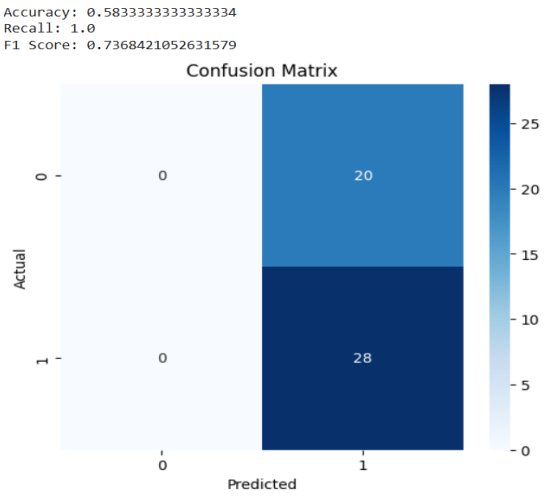
$$F1score = \frac{2*(\Pr ecision + \Re call)}{\Pr ecision + \Re call}$$

Шатастыру матрицасы (Confusion Matrix) – классификатордың жұмысын бағалау үшін қолданылатын матрицалық кесте. Ол әдетте жіктеу алгоритмінің дұрыс және бұрыс нәтижелерінің санын көрсетеді. Ол сондай-ақ бірінші және екінші ретті қателер туралы ақпаратты береді [20].



Сурет 7. Random Forest алгоритмінің нәтижесі

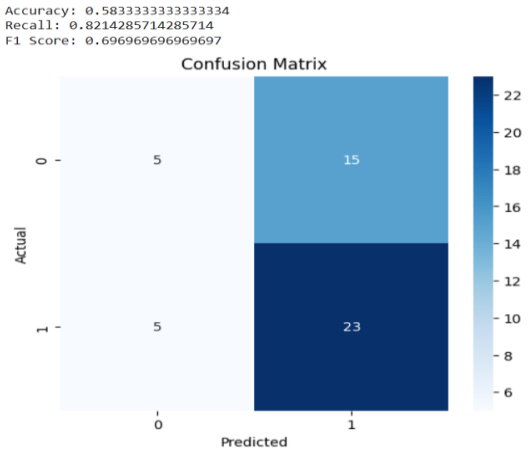
Random Forest нәтижелері: дәлдік жақсы, бірақ II типті қателердің саны да өсті. Бұл модель адамның несие бойынша дефолтқа ұшырайтынын сирек болжайды.



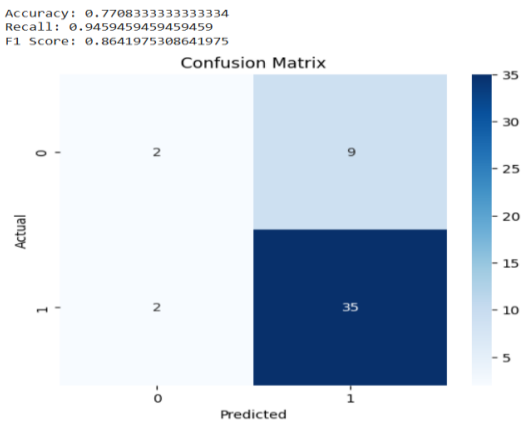
Сурет 8. Векторлық қолдау машинасы (SVM) алгоритмінің нәтижесі

SVM нәтижелері: Бұл модель логистикалық регрессиямен және кездейсоқ орманмен салыстырғанда орташа мәнді орындайды.

Біздің жағдайда бұл алгоритмді қолдану барлық жағынан банктің тәуекелін айтарлықтай азайтпайды.



Сурет 9. KNN алгоритмінің нәтижесі



Сурет 10. Логистикалық регрессия алгоритмінің нәтижесі

Кесте 1. Қолданылған алгоритмдердің нәтижесі

Алгоритмдер	Accuracy	Precision	Recall	F1-score
Кездейсоқ орман (Random Forest)	0.79	0.25	0.9	0.85
Векторлық қолдау машинасы (SVM)	0.58	0.17	1	0.73
k-Nearest Neighbours (KNN)	0.58	0.14	0.82	0.69
Логистикалық регрессия	0.77	0.18	0.94	0.86

4. Қорытынды

Мақалада әртүрлі машина алгоритмдері арқылы несиелік скорингті зерттеу қарастырылған. Зерттеудің негізгі мақсаты деректер жиынтығынан клиенттердің дефолт ықтималдығын болжау үшін тиімді скоринг үлгісін жасау болды. Осы мақсатқа жету үшін үлкен деректер жинағы пайдаланылды, онда клиенттер және олардың несиелік тарихы туралы ақпарат бар. Деректер алдын ала өңделді және үлгіні оқытуға дайындық кезінде тазалау, масштабтау және түрлендіру қадамдарынан өтті. Зерттеуде төрт түрлі машиналық оқыту моделі қолданылды: кездейсоқ орман, векторлық қолдау машинасы (SVM), *K* жақын көршілері (KNN) және логистикалық регрессия. Әрбір модель оқу-жаттығу жиынында оқытылды және сынақ жиынында сынақтан өтті. Тәжірибелер әртүрлі үлгілердің өнімділігін салыстыруға және олардың дәлдігін, еске түсіруін және жіктеу сапасының көрсеткіштеріне негізделген F1 мәнін бағалауға мүмкіндік берді. Нәтижелер кездейсоқ орман және логистикалық регрессия үлгілері ең жақсы дәлдік көрсетті. Random Forest сияқты ансамбльдік әдістерге негізделген модель барлық тексерілген модельдер арасында ең жақсы өнімділікті көрсетті. Ол клиенттердің несие қабілеттілігін болжауда жоғары дәлдік пен тұрақтылықты көрсетті. SVM басқа үлгілермен салыстырғанда нашар өнімділік көрсетті. KNN сынақтан өткендердің арасында ең тиімсіз модель болып шықты. Олардың деректердегі күрделі тәуелділіктерді модельдеу мүмкіндіктері шектеулі және жоғары нәтижелерге жету үшін параметрлерді мұқият таңдауды талап етеді.

References / Әдебиеттер

[1] Steinwart, I., Christmann, A. (2008). Support Vector Machines. *Springer*

[2] Smola, A. (2002). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. *MIT Press*

[3] Altman, N.S. (2002). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician, 46(3), 175–185*

[4] Zhang, Z. (2016). Introduction to Machine Learning: k-Nearest Neighbors. *Annals of Translational Medicine, 4(11), 218*

[5] Kazakstan Respublikasynin Ulttyk Banki. Kazakstannin bank sektory boiynsha taldau zhane statistika. *Retrieved from:* https://nationalbank.kz/ru/news/banks-performance

[6] KR Ulttyk statistika bjurosy. Bank sektory men makroekonomikalyk korsetkishter boiynsha malimetter. *Retrieved from:* https://stat.gov.kz/

[7] Biau, G. (2012). Analysis of a Random Forests Model. *Journal of Machine Learning Research, (13), 1063–1095*

[8] Breiman, L. (2001). Random Forests. *Machine Learning, 45(1), 5–32*

[9] Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X. (2013). Applied Logistic Regression. *Wiley*

[10] Kleinbaum, D.G., Klein M. (2010). Logistic Regression: A Self-Learning Text. *Springer*

[11] Ashimbaev, T.A. (2015). Қазақстанның bank zhүjesi: қalyptasu tarihy zhәne damu perspektivalary. Almaty: Ekonomika

[12] Kenzhegaliev, B. (2020). Қазақстандағы kredittik skoring: mashinalyқ oқytu algoritmderi men әdisteri. *Astana: Nur-Press*

[13] Kazakstan Respublikasynin bank sektoryn taldau: maseleler men damu perspektivalary. (2023). *CyberLeninka*. *Retrieved from:* https://cyberleninka.ru/article/n/analiz-bankovskogo-sektora-respubliki-kazahstan

[14] Ranking.kz. (2023). Kazakstan bankterinin rejtingi zhane karzhylyk korsetkishteri. *Retrieved from:* https://ranking.kz/

[15] Burkov, A. (2019). The Hundred-Page Machine Learning Book. *Retrieved from:* https://themlbook.com/

[16] Bishop, C.M. (2006). Pattern Recognition and Machine Learning. *Springer*

[17] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. *O'Reilly Media*

[18] Sokolova, M., Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45(4), 427-437.* https://doi.org/10.1016/j.ipm.2009.03.002

[19] Willett, P. (2006). The Porter stemming algorithm: then and now. *Program: Electronic Library and Information Systems, 3(40), 219–223*

[20] Porter, M.F. (1980). An algorithm for suffix stripping. *Program, 14(3), 130–137*

Интеллектуалды скоринг және несие тәуекелі

Е. Ыбраев, Ж. Алибиева*

Satbayev University, Алматы, Қазақстан

**Корреспонденция үшін автор: zh.alibiyeva@satbayev.university*

Аңдатпа. Осы мақалада банктік салада несиелік скоринг жүйелерін жетілдіру және несиелік тәуекелдерді басқарудың жаңа әдістері қарастырылады. Ұсынылған жүйе машиналық оқыту әдістерін қолданып, клиенттердің қаржылық тарихын, нарықтық жағдайларды және макроэкономикалық көрсеткіштерді кешенді түрде талдауға негізделеді, бұл несиелік тәуекелді дәлірек бағалауға мүмкіндік береді. Жұмыс барысында әртүрлі модельдердің салыстырмалы талдауы жүргізіліп, олардың тез өзгеретін қаржылық ортада практикалық қолданылуы мен тиімділігі көрсетіледі. Эксперименталды нәтижелер ұсынылған алгоритмдерді қолданудың болжау сапасын айтарлықтай жақсартатынын және қаржылық шығындарды азайтатынын дәлелдейді. Зерттеу банктерде жылдам әрі негізделген басқарушылық шешімдер қабылдауға септігін тигізетін адаптивті аналитикалық құралдарды енгізудің перспективаларын атап көрсетеді.

Негізгі сөздер: *жасанды интеллект, машиналық оқыту, логистикалық регрессия, датасет, деректер қоры, интеграция.*

Интеллектуальный скоринг и кредитный риск

Е. Ыбраев, Ж. Алибиева*

Satbayev University, Алматы, Казахстан

**Автор для корреспонденции: zh.alibiyeva@satbayev.university*

Аннотация. В данной статье исследуются инновационные подходы к совершенствованию систем кредитного скоринга и управления кредитными рисками в банковской сфере. Предлагаемая система использует методы машинного обучения для комплексного анализа финансовой истории клиентов, рыночных условий и макроэкономических показателей, что позволяет проводить более точную оценку кредитного риска. Работа включает сравнительный анализ различных моделей, демонстрируя их практическую применимость и эффективность в условиях быстро меняющейся финансовой среды. Результаты экспериментов показывают, что использование предложенных алгоритмов способствует значительному повышению качества прогнозов и снижению финансовых потерь. Исследование подчеркивает перспективы внедрения адаптивных аналитических инструментов, способствующих принятию оперативных и обоснованных управленческих решений в банках.

Ключевые слова: *искусственный интеллект, машинное обучение, логистическая регрессия, датасет, базы данных, интеграция.*

Received: 13 February 2024
Accepted: 15 June 2024
Available online: 30 June 2024