

4.	ՏՆՏԵՍԱԳԻՏՈՒԹՅՈՒՆ ECONOMICS ЭКОНОМИКА
----	---

**КНН ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՆ ԱԼԳՈՐԻԹՄԻ ԿԻՐԱՌՈՒՄԸ ՖԻԶԻԿԱԿԱՆ
ԱՆՁԱՆՑ ՎԱՐԿԱՅԻՆ ՌԻՍԿԻ ԿԱՌԱՎԱՐՄԱՆ ՀԱՄԱԿԱՐԳՈՒՄ**

ՀՏԴ 336.7, 519.8 004.8

DOI: 10.56246/18294480-2024.17-23

ԱՌԱՔԵԼՅԱՆ ԳԱՌՆԻԿ

*ՀՊՏՀ ԿՐՆԿՐԵՍԱԿԱՆ ԻՆՖՈՐՄԱԿԻԿԱՅԻ և ԿՐԵՂԵԿԱՐՎԱԿԱՆ
համակարգերի ամբիոնի հայցորդ,*

*ՀՊՏՀ ԿՐՆԿՐԵՍԱԿԱՆ ԻՆՖՈՐՄԱԿԻԿԱՅԻ և ԿՐԵՂԵԿԱՐՎԱԿԱՆ
համակարգերի ամբիոնի դասախոս
Էլվինսոն՝ arakelyan0393@mail.ru*

Յուրաքանչյուր վարկապու կազմակերպություն իրականացնում է իր գործունեությունը անկայուն միջավայրում և չունենալով նրա մասին լիարժեք տեղեկություն՝ կարող է կրել զգալի կորուստներ: Ռիսկը հնարավոր կորուստներ կրելու կամ հավելյալ օգուտ ստանալու հավանականության առկայությունն է, որը ուղեկցում է յուրաքանչյուր գործունեությանը: Այն իրենից ներկայացնում է որոշումների կայացման ընթացքում առկա այնպիսի զարգացումների հավանականություն, որոնք կարող են բացասաբար անդրադառնալ ԿՐԵՂԵԿԱՐՎԱԿԱՆ սուբյեկտի շահույթի, կապիտալի, հեղուկաբար ամբողջ գործունեության վրա և հանգեցնել կորուստների: Այսպիսով, առևտրային բանկերի կառավարման առանցքային օղակներից է հանդիսանում ռիսկերի գնահատման և կառավարման արդյունավետ ռազմավարության ընտրությունը:

1997թ. բանկային վերահսկողության բազելյան կոմիտեն իր հիմնական փաստաթղթերում վարկային ռիսկը անվանել է որպես այն հիմնական ֆինանսական ռիսկ, որի հետ են իրենց գործունեության ընթացքում առնչվում ֆինանսական ինստիտուտները:

Աշխարհի ներկայիս տեխնիկական զարգացման մակարդակը թույլ է տալիս ռիսկերի գնահատումը և կառավարումը իրականացնել փարբեր տեխնոլոգիաների միջոցով, մասնավորապես մեքենայական ուսուցման միջոցով:

Սույն հետազոտության շրջանակում իրականացվել է հաճախորդների վարքագծի վերլուծություն կոնկրետ օրինակների հիման վրա, ինչպես նաև մշակել վարկառուների դեֆոլտ կանխատեսելու համար kNN մոդել:

Ուսումնասիրությունները ցույց են տալիս, որ մեքենայական ուսուցման մոդելների կիրառմամբ ֆինանսաբանկային համակարգում կարելի է ստեղծել լուծումներ, որոնք կնվազեցնեն առկա վարկային ռիսկերի ազդեցությունը և կազմակերպությունների ծախսերը:

Քանալի բառեր` վարկային ռիսկ, մեքենայական ուսուցում, KNN (*k*-nearest neighbors), Էվկլիդյան հեռավորություն, *Weights Of Evidence* (WOE).

Ներածություն

Առևտրային բանկերի ավանդական գործառնություններից է հանդիսանում վարկավորումը, որը իր մեջ պարունակում է վարկային ռիսկ, որի կառավարումը բացառիկ տեղ է գրավում բանկային համակարգում՝ նվազեցնելով կորուստ կրելու հավանականությունը: Աշխարհի ներկայիս զարգացման մակարդակը և ֆինանսական ինստիտուտներում գոյացած մեծ ծավալի տվյալները թույլ են տալիս վարկային ռիսկը գնահատել՝ կիրառելով մեքենայական ուսուցման գործիքները, ինչով էլ պայմանավորված է ուսումնասիրման առարկայի արդիականությունը:

Աշխատանքի նպատակն է հետազոտել կոնկրետ օրինակների հիման վրա հաճախորդների վարքագիծը, մշակել վարկառուների դեֆոլտ կանխատեսելու համար kNN մոդել:

KNN (*k*-nearest neighbors) ալգորիթը իրենից ներկայացնում է ոչ պարամետրիկ, վերահսկվող դասակարգիչ, որը դասակարգման կամ կանխատեսման համար կիրառում է տվյալների առանձին կետերի միջև եղած հեռավորությունը¹: Այն իրենից ներկայացնում է մեքենայական ուսուցման պարզ և միաժամանակ հզոր ալգորիթ: KNN մոդելում հիմնական գաղափարը կայանում է նրանում, որ այն կետերը (դիտարկումները), որոնք միմյանցից գտնվում են մոտիկ հեռավորության վրա, նման են իրար, իսկ միմյանցից մեծ հեռավորության վրա գտնվող կետերը (դիտարկումները) տարբերվում են իրարից: Այս մոդելի շրջանակում հեռավորությունը կարելի է չափել տարբեր մեթոդներով, հաճախ կիրառվում է Էվկլիդյան հեռավորությունը, ինչի հաշվարկման հիմքում ընկած է Պյութագորասի թեորեմը²: Ստորև Էվկլիդյան հեռավորությունը չափելու համար նախատեսված բանաձևը.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Վերոնշյալ մոդելում նախապես ընտրվում է մոտակա հարևանների քանակ (*k*), որը դասակարգչում առանցքային դեր է խաղում: Պատկերացնենք, որ առկա է մոդել, որը դասակարգում է օբյեկտները կարմիր կամ կապույտ գույների կատեգորիաների:

¹ <https://www.ibm.com/topics/knn>

² <https://medium.com/@luigi.fiori.lf0303/distance-metrics-and-k-nearest-neighbor-knn-1b840969c0f4>

Ենթադրենք, որ մոտակա հարևանների քանակը ընտրված է 5, ապա մոդելը դասակարգմանը նախատեսված դեռ անհայտ օբյեկտի համար հաշվարկելու է իրենից բոլոր կետերի միջև հեռավորությունները և դրանցից ընտրելու է առավել մոտիկ 5 դիտարկումները: Այնուհետև, անհայտ օբյեկտը դասակարգվելու է այն կատեգորիային (կարմիր, կապույտ), որից քանակությամբ շատ է լինելու այդ ընտրված 5 դիտարկումներում:

Վերոնշյալ ալգորիթմում առանցքային դեր է խաղում օպտիմալ մոտակա հարևանների քանակի ընտրությունը:

Հիմնական մաս

Հետազոտության ընթացքում ուսումնասիրվել են թեմային առնչվող տարբեր հոդվածներ, թեզեր, գրքեր և այլն: Դիտարկենք օրինակներ:

Բոբկով Ա.Պ. կողմից գրված «Использование методов машинного обучения для оценки рисков при внедрении нового кредитного продукта» հոդվածում դիտարկվում է անձանց վարկունակության գնահատման հարցերը վարկային սքորինգի կիրառմամբ: Աշխատանքում սքորինգի մոդելը ստեղծվում է կորպորատիվ պահուցում պահվող հաճախորդների տվյալների՝ ընտրված ենթաբազմության հիման վրա¹: Սի Շի, Ռիտա Ցե, Ռուման Լուո, Ստեֆանո Դ'Ադդոնա և Ջովանի Պաու կողմից գրված «Machine learning-driven credit risk: a systemic review» հոդվածում ուսումնասիրվել է վերջին 8 տարիներին գրված թվով 76 հոդված, որտեղ կիրառվել են վիճակագրության մեթոդները՝ մեքենայական ուսուցումը և խորը ուսուցումը վարկային ռիսկի հետ կապված խնդիրները լուծելու համար: Հեղինակները առաջարկել են վարկային ռիսկի ալգորիթմների դասակարգման նոր մեթոդ և քննարկել են տվյալների դիսբալանսի խնդիրը: Որպես արդյունք՝ հեղինակները նշում են, որ խորը ուսուցման մոդելների գերակշիռ մասը առավել արդյունավետ են, քան դասական մեքենայական ուսուցման և վիճակագրական ալգորիթմները, ինչպես նաև ensemble մեթոդները առավել ճշգրիտ են, քան միայնակ մոդելները²:

Այս հետազոտության շրջանակում kNN մոդելի մշակման համար հիմք են ընդունվում հաճախորդների վարկերի իրական տվյալներ «Յունիբանկ» ԲԲԸ-ի օրինակով (4603 դիտարկում): Առկա բազան բաղկացած է թվով 13 փոփոխականներից՝ վարկառուի տարիքը վարկը վերցնելու պահին (20-ից մինչև 65 տարեկան), վարկառուի սեռը (իգական, արական), վարկառուի ամուսնական կարգավիճակը (միայնակ, ամուսնացած, բաժանված, այրի), վարկառուի կրթության աստիճանը (գիտական աստիճան, բարձրագույն կրթություն, միջին մասնագիտական կրթություն, դպրոցական կրթություն), սեփականության առկայությունը (անշարժ գույքի առկայությունը, շարժական գույքի առկայությունը, անշարժ և շարժական գույքերի առկայու-

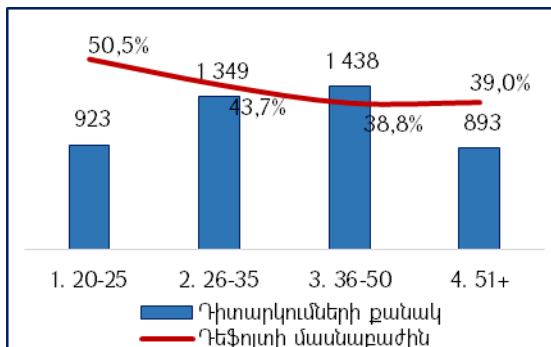
¹ <https://cyberleninka.ru/article/n/ispolzovanie-metodov-mashinnogo-obucheniya-dlya-otsenki-riskov-pri-vnedrenii-novogo-kreditnogo-produkta/viewer>

² <https://link.springer.com/article/10.1007/s00521-022-07472-2>

թյունը, սեփականության բացակայությունը), պայմանագրային գումարը (30,000-ից մինչև 2,300,000 ՀՀ դրամ), վերջին 12 ամիսների ընթացքում ժամկետանց օրերի քանակը (0-ից մինչև 1600), ուշացումների քանակը (0-ից մինչև 67), ռիսկի դասերի փոփոխությունների քանակը (0-ից մինչև 28), վարկային բեռը (0-ից մինչև 2,984,303 ՀՀ դրամ), վարկային պատմության երկարությունը (0-ից մինչև 488 օր), առավելագույն մարած վարկը (0-ից մինչև 25,131,048 ՀՀ դրամ), դեֆոլտը (0-ոչ, 1- այո):

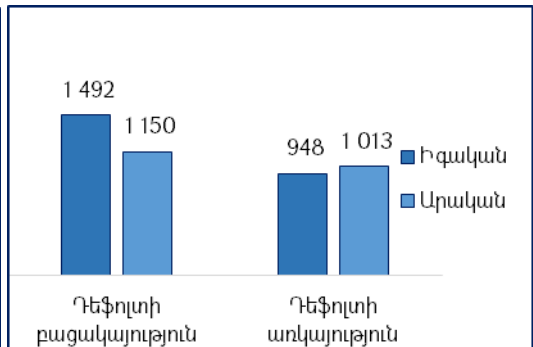
Աշխատանքի շրջանակում տվյալների վերլուծության, մշակման և մոդելավորման համար կիրառվել են Excel ծրագիրը, Python ծրագրավորման լեզվի Pandas, Matplotlib, Seaborn, Scikit-learn գրադարանները:

Տվյալների համար իրականացվել է կորելյացիոն վերլուծություն, որտեղից բացահայտվել է, որ առկա է ուժեղ կապ վարկային բեռի և վարկային պատմության երկարության (0,5), ռիսկի դասերի փոփոխությունների քանակի և ուշացումների քանակի (0.57) միջև: Իրականացվել են տարբեր փոփոխականների վերլուծություններ: Ստորև որոշ փոփոխականների մասով գրաֆիկական վերլուծություն (գծապատկեր 1-4)



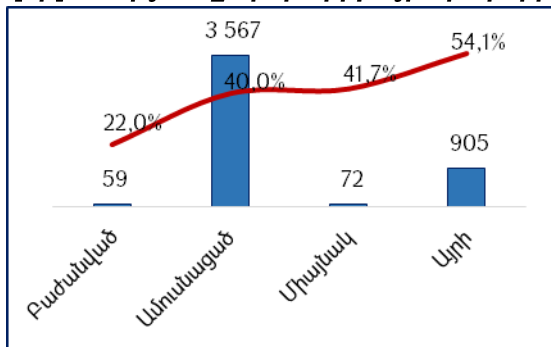
Գծապատկեր 1. Տվյալների

վերլուծություն ըստ տարիքային խմբերի:



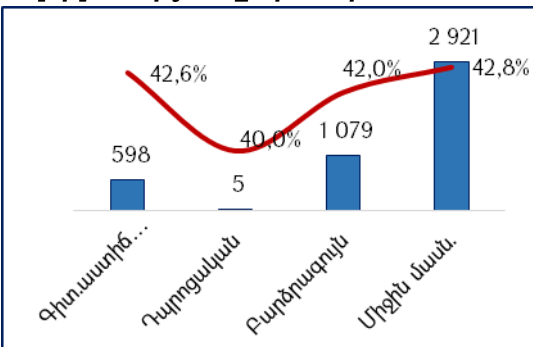
Գծապատկեր 2. Տվյալների

վերլուծություն ըստ սեռի:



Գծապատկեր 3. Տվյալների

վերլուծություն ըստ ամուսնական կարգավիճակի:



Գծապատկեր 4.

Տվյալների վերլուծություն ըստ կրթության աստիճանի:

Մոդելավորման նպատակով առկա տվյալները անհրաժեշտ է խմբավորել և բոլոր խմբերի համար կատարել ստանդարտացում: Ստանդարտացման հիմնական

նպատակն է բերել բոլոր փոփոխականները մեկ ընդհանուր տեսքի: Գործնականում գոյություն ունեն ստանդարտացման տարբեր մեթոդներ: Սույն աշխատանքում կիրառվում է Weights Of Evidence (WOE) մեթոդը, քանի որ այն լայնորեն տարածված և հասկանալի է¹: Խմբավորված գործակցի յուրաքանչյուր խմբի համար WOE հաշվարկվում է հետևյալ բանաձևով

$$WOE = \ln \frac{\text{Percentage of good in the class}}{\text{Percentage of bad in the class}},$$

որտեղ ln – բնական լոգարիթմն է:

Ստորև աղյուսակ 1-ում ներկայացված են խմբավորված և ստանդարտացված փոփոխականները:

Աղյուսակ 1. Խմբավորված և ստանդարտացված փոփոխականներ (*-ով ներկայացված են հազ. ՀՀ դրամով):

Խումբ	Ընդհամենը*	WOE	Խումբ	Ընդհամենը*	WOE
Սեռ			Վերջին 12 ամիսների ընթացքում ժամկետանց օրերի քանակ		
Իգական	612,678	0,169	1. 0-30	1,155,316	0,024
Արական	557,540	-0,182	2. 31+	14,902	-2,520
Կրթություն			Սեփականության առկայություն		
Գիտական աստիճան	155,561	0,068	Անշարժ գույքի առկայություն	70,760	0,047
Բարձրագույն	291,127	0,053	Շարժական գույքի առկայություն	280,834	0,052
Միջին մասնագիտական	722,190	-0,037	Անշարժ և շարժական գույքի առկայություն	59,682	0,145
Միջնակարգ	1,339	0,688	Բացակայություն	758,940	-0,035
Ընտանեկան կարգավիճակ			Վարկառուի տարիք		
Ամուսնացած	913,540	0,093	1. 20-25	210,717	-0,336
Այրի	225,660	-0,413	2. 26-35	350,813	-0,090
Բաժանված	13,085	0,995	3. 36-50	370,738	0,162
Միայնակ	17,931	-0,139	4. 51+	237,949	0,185
Ուշացումների քանակ			Ռիսկի դասերի փոփոխությունների քանակ		
1. 0	540,573	0,194	1. 0	1,038,196	-0,044

¹ <https://medium.com/mlearning-ai/weight-of-evidence-woe-and-information-value-iv-how-to-use-it-in-eda-and-model-building-3b3b98efe0e8>

2. 1	629,644	-0,164	2. 2+	132,021	0,360
Վարկային բեռ			Վարկային պատմության երկարություն		
1. 0	317,685	0,729	1. 0-270	1,156,563	-0,002
2. 1-300,000	207,519	0,001	2. 271-365	10,143	0,169
3. 300,001+	645,013	-0,331	3. 366+	3,512	0,254
Պայմանագրային գումար			Առավելագույն մարած վարկ		
1. Մինչև 200,000	520,209	0,179	1. 0-350,000	810,315	0,008
2. 200,001-400,000	418,168	0,127	2. 350,001+	359,903	-0,017
3. 400,001+	231,840	-0,628			

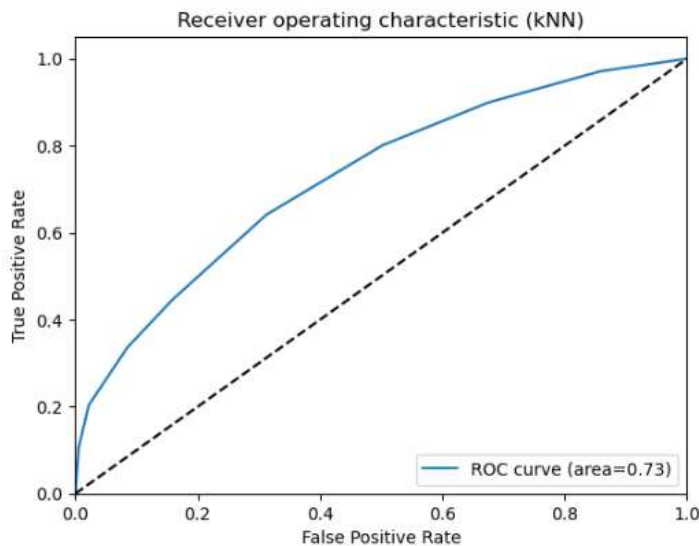
Սույն աշխատանքում մեքենայական ուսուցման համար որպես անկախ փոփոխականներ հիմք են ընդունվում յուրաքանչյուր խմբի համար հաշվարկված WOE արժեքները, իսկ որպես կախյալ փոփոխական, որը անհրաժեշտ է կանխատեսել՝ “Default” բնութագրիչը, որը կարող է լինել 0 (ոչ դեֆոլտ) կամ 1 (դեֆոլտ): Մոդելի լավագույն արդյունք ստանալու, ինչպես նաև գերուսուցման խնդիրը լուծելու համար պետք է որոնել համապատասխան պարամետրերը քրոս-վալիդացիայի և GridSearchCV գործիքի միջոցով (գործիքի միջոցով օպտիմալ է ճանաչվել մոտակա 8 հարևանը և էվկլիդյան հեռավորությունը): Ուսուցման համար տվյալները բաժանվել են երկու խմբի՝ ուսուցման (x_{train} , y_{train}) և թեստավորման (x_{test} , y_{test})՝ 65% և 35% հարաբերակցությամբ: Վերոնշյալ գործողությունները կատարելուց հետո իրականացվել է մոդելի ուսուցում x_{train} և y_{train} տվյալներով, որի հիման վրա բացահայտվել են դրա լավագույն պարամետրերը: Այդ պարամետրերով ստեղծված մոդելը պահպանվել է և այդ մոդելի մասով հաշվարկվել են որակի ցուցանիշ sklearn գրադարանի scor մեթոդով՝ ուսուցման համար նախատեսված տվյալների հիման վրա 73,9%, իսկ թեստավորման համար նախատեսված տվյալների հիման վրա՝ 68,0%: Թեստային տվյալների հիման վրա հաշվարկվել է նաև մոդելի որակի ճշգրտության գործակիցը (Precision՝ 0,66) և մոդելի որակի այն գործակիցը, որը ցույց է տալիս ճիշտ կանխատեսված պատասխանների մասնաբաժինը (Recall՝ 0,45): Այնուհետև յուրաքանչյուր մոդելի համար հաշվարկվել են F1 score (0,53), որը գաղափարապես միավորում է վերոնշյալ ցուցանիշների ինֆորմացիան¹:

Կառուցվել է նաև մոդելի համար ROC կորը, որը թույլ է տալիս գնահատել մոդելի որակը՝ ցույց տալով հարաբերակցությունը ճիշտ դասակարգված և սխալ դասակարգված օբյեկտների միջև (գծապատկեր 5): ROC-ի քանակական մեկնաբանությունը տալիս է AUC (Area Under Curve) ցուցիչը՝ մակերեսը, որը սահմանափակված է ROC կորով և կեղծ դրական դասակարգումների մասնաբաժնի առանցքով: Որքան բարձր է AUC ցուցանիշը, այնքան ավելի որակյալ է դասակարգիչը, մինչդեռ 0,5

¹ Н. Шакла, «Машинное обучение & TensorFlow», – Питер, 2019 г., стр. 118

արժեքը ցույց է տալիս ընտրված դասակարգման մեթոդի ոչ պիտանելիությունը: 0,5-ից փոքր արժեքը ցույց է տալիս, որ դասակարգիչը գործում է ճիշտ հակառակը¹:

Գծապատկեր 5.. Մոդելի ROC կորեր:



Թեստային տվյալների հիման վրա միայն դեֆոլտ գնացած հաճախորդներին տրված ընդհանուր վարկերի ծավալը կազմում է 178 մլն.դրամ, որից 88,6 մլն.դրամ մշակված kNN մոդելը ճիշտ է կանխատեսել:

Ամփոփիչ մաս

Աշխատանքի արդյունքից կարելի է եզրակացնել, որ մեքենայական ուսուցման մոդելների կիրառումը (մասնավորապես տվյալ օրինակի հիման վրա kNN մոդելի կիրառումը) ֆինանսաբանկային համակարգում գործնականում արդյունավետ է և կարող է նվազեցնել առկա վարկային ռիսկերի ազդեցությունը, կազմակերպությունների ծախսերը (տվյալ օրինակով՝ մոտ 88,6 մլն. դրամ): Ելնելով վերոնշյալից և ներկայիս արագ տեխնիկական զարգացումից, ինչպես նաև բանկային համակարգում գոյություն ունեցող թեժ մրցակցության պայմաններից՝ առաջարկվում է ֆինանսական հաստատություններին ներդնել մեքենայական ուսուցման տեխնոլոգիաները վարկավորման, այդ թվում վարկային ռիսկի կառավարման գործընթացում:

Օգտագործված գրականության ցանկ

1. <https://www.ibm.com/topics/knn>

¹ <https://www.ibm.com/docs/ru/spss-statistics/beta?topic=features-roc-analysis>

2. <https://medium.com/@luigi.fiori.lf0303/distance-metrics-and-k-nearest-neighbor-knn-1b840969c0f4>
3. <https://cyberleninka.ru/article/n/ispolzovanie-metodov-mashinnogo-obucheniya-dlya-otsenki-riskov-pri-vnedrenii-novogo-kreditnogo-produkta/viewer>
4. <https://link.springer.com/article/10.1007/s00521-022-07472-2>
5. <https://medium.com/mlearning-ai/weight-of-evidence-woe-and-information-value-iv-how-to-use-it-in-eda-and-model-building-3b3b98efe0e8>
6. Н. Шакла, «Машинное обучение & TensorFlow», – Питер, 2019 г., стр. 336
7. <https://www.ibm.com/docs/ru/spss-statistics/beta?topic=features-roc-analysis>

USE OF KNN MACHINE LEARNING ALGORITHM IN PHYSICAL ENTITY CREDIT RISK MANAGEMENT SYSTEM

ARAKELYAN GARNIK

*PhD Student and Lecturer of Chair of Economic Computer Science
and Information Systems,
Armenian State University of Economics
e-mail: arakelyan0393@mail.ru*

Any credit institution operates in an unstable environment and without complete information, it can suffer significant losses. Risk is the probability of incurring possible losses or obtaining additional benefits that accompanies any activity. It represents the probability of such a development of events whereupon making decisions may negatively affect the profit, capital, and therefore the entire activity of the business entity thus leading to losses. Therefore, the choice of an effective risk assessment and management strategy is one of the key links in the management of commercial banks.

In 1997, in its main documents the Basel Committee on Banking Supervision called credit risk the main financial risk faced by financial institutions in the process of their activities.

The current level of technical development in the world allows risk assessment and management through various technologies, in particular, through machine learning.

As part of this study, an analysis of customer behavior was carried out using specific examples, and a KNN model was developed to predict the borrower's default.

The results of the work show that with the help of machine learning models in the financial and banking system, solutions can be created that will reduce the impact of existing credit risks and expenses of organizations.

Keywords: *credit risk, machine learning, KNN (k-nearest neighbors), Euclidean distance, Weights of Evidence (WOE)*

ИСПОЛЬЗОВАНИЕ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ KNN В СИСТЕМЕ УПРАВЛЕНИЯ КРЕДИТНЫМИ РИСКАМИ ФИЗИЧЕСКИХ ЛИЦ

АРАКЕЛЯН ГАРНИК

*Соискатель кафедры экономической информатики
и информационных систем АГЭУ*

*Преподаватель кафедры экономической информатики
и информационных систем АГЭУ*

электронная почта: arakelyan0393@mail.ru

Любое кредитное учреждение осуществляет свою деятельность в нестабильной среде и не имея полной информации о ней может понести значительные потери. Риск – наличие вероятности понесения возможных потерь или получения дополнительных выгод, которые сопровождают любую деятельность. Он представляет собой вероятность такого развития событий при принятии решений, которое может негативно отразиться на прибыли, капитале, а следовательно, и на всей деятельности хозяйствующего субъекта и привести к убыткам. Таким образом, выбор эффективной стратегии оценки и управления рисками является одним из ключевых звеньев управления коммерческими банками.

В 1997 году Базельский комитет по банковскому надзору в своих основных документах назвал кредитный риск основным финансовым риском, с которым сталкиваются финансовые институты в процессе своей деятельности.

Современный уровень технического развития в мире позволяет осуществлять оценку и управление рисками посредством различных технологий, в частности посредством машинного обучения.

В рамках данного исследования был проведен анализ поведения клиентов на конкретных примерах, а также разработана модель kNN для прогнозирования дефолта заемщика.

Результаты работы показывают, что с помощью моделей машинного обучения в финансово-банковской системе можно создавать решения, которые позволят снизить влияние существующих кредитных рисков и затрат организаций.

Ключевые слова: кредитный риск, машинное обучение, KNN (*k*-nearest neighbors), Евклидово расстояние, Weights Of Evidence (WOE)

Հոդվածը ներկայացվել է խմբագրական խորհուրդ 01.03.2024թ.:

Հոդվածը գրախոսվել է 17.03.2024թ.:

Ընդունվել է տպագրության 30.11.2024թ.: