

ՀՏԴ 336.7Տնտեսագիտություն**Արմեն ՂԱԶԱՐՅԱՆ****ՀՊՏՀ տնտեսական ինֆորմատիկայի և տեղեկատվական համակարգերի ամբիոնի դոցենտ, տ.գ.թ.****Էլ. հասցե: armenghazaryan@mail.ru**

ՎԱՐԿԱՑԻՆ ՍՔՈՐԻՆԳԻ ԵՎ ԱՆՁԻ ՆՈՒՅՆԱԿԱՆԱՑՄԱՆ ՀԱՎԵԼՎԱԾՆԵՐԻ ՆԱԽԱԳԾՈՒՄԸ ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՄԲ

Հոդվածում առաջարկվել է բանկի հաճախորդների վարկունակության կանխատեսման ալգորիթմ՝ ինչպես մեքենայական ուսուցման առանձին մոդելների, այնպես էլ դրանց համադրության հիման վրա: Բացի դրանից, մշակվել են վարկային սրորինգի և անձի նույնականացման հավելվածներ՝ մեքենայական ուսուցման կիրառմամբ:

Բանալի բառեր՝ հաճախորդների վարկունակության, մեքենայական ուսուցման մեթոդներ, արհեստական բանականություն, լոգիստիկ ռեգրեսիա, արհեստական նեյրոցանց:

А. Казарян ПРОГНОЗ КРЕДИТОСПОСОБНОСТИ КЛИЕНТОВ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

В статье был предложен алгоритм прогнозирования кредитоспособности клиентов банка как с использованием отдельных моделей машинного обучения, так и с применением их комбинации. Кроме того, разработаны приложения для кредитного скоринга и идентификации личности с применением машинного обучения:

Ключевые слова: кредитоспособность клиентов, методы машинного обучения, искусственный интеллект, логистическая регрессия, искусственный нейросеть.

A. Ghazaryan DESIGNING APPLICATIONS FOR CREDIT SCORING AND IDENTITY IDENTIFICATION WITH MACHINE LEARNING

The article proposes an algorithm for predicting the creditworthiness of bank customers both using separate machine learning models and using a combination of them. In addition, applications for credit scoring and identity identification using machine learning have been developed:

Key words: customer creditworthiness, machine learning methods, artificial intelligence, logistic regression, artificial neural network:

Մեր աշխատանքի նպատակն է կատարելագործելը հաճախորդների վարկունակության կանխատեսման մեթոդաբանությունն, կայացնել վարկի տրամադրման վերաբերյալ և բանկում ներդնել հաճախորդի բիոմետրիկ նույնականացման գործընթաց՝ կիրառելով արհեստական բանականության տեխնոլոգիաները:

Վարկային ծառայությունների ոլորտում, շուկայական մրցակցության աճի հետ կապված, վարկունակության գնահատման առանձին տարրերի մշակումը և վարկային ռիսկի ճշգրիտ որոշումը համարվում են արդիական խնդիր: Բացի դրանից, բանկերում առցանց ծառայությունների ներդրմանը զուգընթաց նվազում է հաճախորդի բանկ այցելելու հաճախականությունը, նպաստում սպասարկման ծախսերի կրճատմանը, ծառայությունների որակի բարելավմանը: Սակայն առցանց գործարքների մեծ մասի համար (քարտի կամ վարկի համար դիմելիս, հաշիվ բացելիս և այլն) հաճախորդից պահանջվում է անցնել բիոմետրիկ նույնականացման գործընթաց: Այս խնդիրը լուծելու համար այսօր բազմաթիվ միջազգային բանկեր արդեն կիրառում են արհեստական բանականության տեխնոլոգիաները:

Մեր կողմից մշակված դեմքի ճանաչման մոդելը հնարավորություն է տալիս համեմատել հաճախորդի՝ այդ պահին արված նկարը անձնագրի նկարի հետ: Մոդելի կառուցման համար օգտագործվել են Python ծրագրավորման լեզվի dlib, openCV գրադարանները: Ստեղծված վեբ հավելվածի միջոցով օգտատերը նախ կարողանում է իր համակարգչի տեսախցիկով ինքնանկար անել, այնուհետև նաև վերբեռնել իր լուսանկարը: olib գրադարանի դեմքի ճանաչման գործիքը մարդու նկարը տեղավորում է 128 չափանի վեկտորական տարածությունում, որտեղ նույն անձի լուսանկարները միմյանց մոտ են գտնվում, իսկ տարբեր անձանցը՝ հեռու¹: Այնուհետև հաշվարկվում է էվկլիդեսյան հեռավորությունը, եթե այն փոքր է 0,1-ից, կանխատեսումը դրական արդյունք է տալիս, հակառակ դեպքում՝ բացասական:

Ստորև մանրամասն քննենք նաև մեր կողմից ստեղծված մեկ այլ հավելված՝ մեքենայական ուսուցման ալգորիթմների վրա կառուցված վարկառուի վարկունակության գնահատման մոդելը:

Վարկային սքորինգը արժեքավոր գործիք է բարելավելու համար ֆինանսական ռեսուրսների արդյունավետությունն ու հասանելիությունը (ֆիզիկական անձանց, փոքր և միջին ձեռնարկություններին): Լավ մշակված սքորինգային մոդելն օգնում է ֆինանսական կառույցներին՝ նվազեցնելու ֆինանսական ռիսկերը և ճիշտ բաշխելու ֆինանսական ռեսուրսները:

Վարկային սքորինգի համար կիրառվող մեթոդները վերջին տարիներին ինչպես քանակով, այնպես էլ բարդության աստիճանով մեծ աճ են գրանցել: Դրանք ներառում են ինչպես ավանդական վիճակագրական, այնպես էլ նորարարական մեթոդներ, ինչպիսիք են՝ արհեստական բանականությունը, ներառյալ մեքենայական ուսուցումը (պատահական անտառ, gradient boosting և

¹http://dlib.net/face_recognition.py.html

խորը նեյրոնային ցանցեր)¹: Ի տարբերություն ավանդական մեթոդների՝ մեքենայական ուսուցման վրա հիմնված վարկային սքորինգի մոդելներն իրականացնում են տվյալների մանրագնին գնահատում և ներառում այնպիսի տվյալներ, որոնք հաճախ կարող են ավանդական մեթոդի ժամանակ թվալ անկարևոր և ընդհանրապես չներառվել: Նորարարական մեթոդները նաև ինքնակատարելագործվելու հնարավորություն ունեն տվյալների բազա նոր տվյալների մուտքագրման դեպքում. հատկանիշ, որը բացակայում է ավանդական մեթոդների դեպքում:

Ստորև կներկայացնենք մեքենայական ուսուցման վրա հիմնված վարկային սքորինգի մոդել: Որպես ուսուցման հիմնական ձև՝ օգտագործվել է անսամբլ ուսուցումը (Ensemble learning): Որպես գործիք կիրառվել է Python ծրագրավորման լեզուն, մասնավորապես՝ հետևյալ հիմնական գրադարանները՝ sklearn, imblearn, pandas, numpy: Վարկային տվյալները (493 520), որոնք օգտագործվել են մոդելի ուսուցման համար, վերցված են բաց հանրային ռեսուրսից՝ Kaggle ինտերնետային կայքից²: Մշակվել է վարկային սքորինգի մոդել, յուրաքանչյուր մուտքագրվող վարկատուի տվյալներին՝ համապատասխան կայացվում է որոշում վարկի տրամադրման վերաբերյալ՝ տալով դրական կամ բացասական եզրակացություն: Արդյունքները ցուցադրելու համար մշակվել է վեբ հավելված (Web application):

Արհեստական բանականությունը ներառում է բազմաթիվ տեսություններ, մեթոդներ և տեխնոլոգիաներ: Մեքենայական ուսուցումն արհեստական բանականության ենթաբաժին է: Այն ավտոմատացնում է վերլուծական մոդելի կառուցումը: Մեքենայական ուսուցումը օգտագործում է մեթոդներ նեյրոնային ցանցերից, վիճակագրությունից, գործառնությունների հետազոտությունից և ֆիզիկայից՝ գտնելու համար տվյալներում թաքնված տեղեկատվությունը/ճշմարտությունը՝ առանց հստակ ծրագրավորված լինելու, թե որտեղ նայել, փնտրել կամ ինչ եզրակացնել³:

Անսամբլ ուսուցումը մեքենայական ուսուցման պարադիգմ է, որտեղ բազմաթիվ մոդելներ (որոնք հաճախ անվանում են «թույլ սովորողներ») վերապատրաստվում են նույն խնդիրը լուծելու համար և համատեղվում՝ ավելի լավ արդյունքներ ստանալու համար: Հիմնական վարկածն այն է, որ երբ թույլ մոդելները ճիշտ են զուգորդվում, ապա կարող են բերել առավել ճշգրիտ և /կամ ուժեղ մոդելների: Այն հիմնված է այն ենթադրության վրա, որ յուրաքանչյուր մոդել դիտարկում է տվյալները տարբեր տեսանկյուններից՝ վեր հանելով ճշմարտության միայն մի մասը: Այնուհետև միավորելով լավ կատարողական ունեցող մոդելները, որոնք վարժեցվել են միմյանցից անկախ, կբերեն ճշմարտության մեծ մասի բացահայտմանը առավել քան մեկ մոդելը: Ուստի, սա

¹ The World Bank Group, *Credit Scoring Approaches Guidelines*, 2019

<http://pubdocs.worldbank.org/en/935891585869698451/CREDIT-SCORING-APPROACHES-GUIDELINES-FINAL-WEB.pdf>

² <https://www.kaggle.com/wordsforthewise/lending-club>

³ SAS, 2019, https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html

կիանգեցնի առավել ճշգրիտ կանխատեսումների և առավել քիչ սխալների: Մոդելների քանակի ավելացմանը զուգահեռ աճում է անսամբլ մոդելի կատարողականը: Եվ պետք է ընտրել մոդելներն այնքան տարբեր, ինչքան հնարավոր է: Սա փոքրացնում է մոդելների միջև կոռելյացիան և հանգեցնում անսամբլ մոդելի կատարողականի բարձրացմանը:

Տարբեր անսամբլ մեթոդները կառուցում են մոդելների համախմբությունը տարբեր կերպ: Ստորև ներկայացնենք հաճախ կիրառվող մեթոդները.

- Բլենդինգ (Blending)՝ բոլոր մոդելների կանխատեսումների միջինացում¹,
- Բագգինգ (Bagging)՝ տվյալների վարժեցվող հավաքածուի պատահական ենթաբազմությունների վրա միմյանցից անկախ մոդելների կառուցում, դրանց անհատական կանխատեսումների միջինացում՝ որպես վերջնական կանխատեսում,
- Բուսթինգ (Boosting)՝ մոդելների հաջորդական կառուցում այնպես, որ հաջորդ մոդելը սովորում է իր նախորդ մոդելից: Վերաուսուցման սխալից խուսափելու համար յուրաքանչյուր մոդել սովորում է նախորդ բոլոր մոդելների արդյունքների հիմքի վրա²,
- Սթեքինգ (Stacking)՝ k բազային սովորողների կառուցում, դրանց արդյունքի տեղավորում մոդելում՝ վերջնական արդյունքի կանխատեսման համար³:

Մոդելավորման համար կիրառվել է անսամբլ մոդելում բագգինգ մեթոդի երկու դասակարգիչ՝ Bagging Classifier-ը և RandomForestClassifier-ը: BaggingClassifier-ը՝ որպես միասնական մետագնահատիչ, ընդունում է որպես մուտքային տվյալ օգտատիրոջ կողմից սահմանված բազային գնահատիչը (base estimator), ինչպես նաև այլ պարամետրեր, որոնցով սահմանվում է պատահական ենթաբազմություններ կազմելու ռազմավարությունը⁴: Որպես բազային գնահատիչ պարամետրի արժեք՝ տրվում է RandomForestClassifier-ը: Պատահական անտառում անսամբլի յուրաքանչյուր ծառ կառուցված է վարժեցման հավաքածուի ենթահավաքածուի վրա: Հիմքում ընկած գաղափարն է աճեցնել մոտավորապես անկախ ծառերի մոդելների շարք, որոնք համատեղ ավելի լավ են աշխատում, քան մեկ ծառի մոդելը: Դասակարգման լուծումները և ռեգրեսիայի արժեքները այնուհետև տրվում են համապատասխանաբար բոլոր առանձին ծառերի «ձայների» մեծամասնությամբ կամ միջին կանխատեսումով:

Մոդելի ուսուցման համար օգտագործվել են տվյալների 80%-ը, իսկ վարժեցման՝ 20%-ը: Տվյալների հավաքածուն ներառում է հետևյալ հատկանիշները (features).

¹ <http://www.machinelearning.ru/wiki/images/5/56/Guschin2015Stacking.pdf>

² <https://neurohive.io/ru/osnovy-data-science/ansamblevye-metody-begging-busting-i-steking/>

³ <https://dyakonov.org/2017/03/10/c%D1%82%D0%B5%D0%BA%D0%B8%D0%BD%D0%B3-stacking-%D0%B8-%D0%B1%D0%BB%D0%B5%D0%BD%D0%B4%D0%B8%D0%BD%D0%B3-blending/>

⁴ <https://scikit-learn.org/stable/modules/ensemble.html>

- **Նպատակ՝** վարկի նպատակը (վարկային քարտ, կոնսումիդացիա, վերանորոգում, անշարժ գույքի ձեռքբերում, շարժական գույք, խոշոր առք, էներգետիկ, բիզնես վարկ, առողջական, հանգիստ, հարսանեկան և այլն),
- **Տոկոսադրույք՝** վարկի տոկոսադրույքը,
- **Ամսական վճարումների մեծություն՝** վարկառուի կողմից վարկի հաստատումից հետո ամսական տրվող վճարների մեծությունը,
- **Տարեկան եկամուտ՝** վարկառուի տարեկան եկամտի մեծությունը,
- **Վարկ/եկամուտ հարաբերակցություն,**
- **Վարկային գնահատական՝** վարկառուի վարկային գնահատականը ըստ FICO համակարգի¹,
- **Վարկառուի հաշվեկշիռ,**
- **Վարկառուի հաշվեկշռի օգտագործման գործակիցը,**
- **Վերջին 6 ամիսների վարկառուի՝ վարկատուներին դիմելու քանակ,**
- **Վերջին 2 տարիներին 30 և ավելի օր ժամկետանց վարկեր ունենալու դեպքերի քանակ,**
- **Վարկային գրանցամատյանի մեջ բացասական նշումների քանակ:**

← → ↻ ⓘ localhost:5000

Մուտքագրեք հաճախորդի տվյալները

Նպատակ

Տոկոսադրույք (%)

Ամսական վճարումների մեծություն (\$)

Տարեկան եկամուտ (\$)

Վարկ/եկամուտ հարաբերակցություն

FICO համակարգում վարկային գնահատական

Վարկառուի հաշվեկշիռ

Վարկառուի հաշվեկշռի օգտագործման գործակից

Վերջին 6 ամիսների վարկատուներին դիմելու քանակ

Վերջին 2 տարիներին 30 և ավելի օր ժամկետանց վարկեր ունենալու դեպքերի քանակ

Վարկային գրանցամատյանի մեջ բացասական նշումների քանակ

Ստեղծված վեր հավելվածում, որը պարզ ինտերֆեյս է ստեղծում օգտագործողի համար, մուտքագրելով վարկառուի մասին տվյալները, որոնք համապատասխանում են վերը նշված տվյալների հավաքածուի

¹ՖՄՅԳՆ Սքորը (FICO® Score) սքորինգային համակարգ է, որը, վիճակագրական ուսումնասիրության և վերլուծության ենթարկելով հաճախորդի վարկային պատմությունը, տալիս է վարկառուի վարկային ռիսկի (վարկարժանության) թվային գնահատականը՝ որպես սպառողական կրեդիտավորման ռիսկի մակարդակի չափորոշման առանցքային գործիք։ Հայաստանում այն մշակվել է «ԱՔՌԱ Քրեդիտ Ռեփորթինգ» ՓԲ և «Ֆեյր Այգաք Սերվիսիզ Լիմիթիդ» ընկերությունների կողմից։

https://acra.am/?page_id=609&lang=hy

հատկանիշներին, օգտատերը ստանում է պատասխան վարկի տրամադրման մասին՝ դրական կամ բացասական: Տվյալների մուտքագրումից հետո **կանխատեսել** ստեղծի սեղմամբ տվյալները փոխանցվում են Front End-ից դեպի Back End, որտեղ կանչվում է համապատասխան մոդելի կանխատեսման ֆունկցիան և որպես արդյունք վերադարձվում է արժեք՝ 1 կամ 0: 1 արժեքի դեպքում օգտատերը տեսնում է կանխատեսման դրական արդյունք, 0 արժեքի դեպքում՝ բացասական: Բոլոր վարկառուների տվյալները պահպանվում են տվյալների բազայում:

Այս ալգորիթմը թույլ է տալիս օգտագործել ինչպես առանձին մոդելներ, այնպես էլ դրանց համակցությունները: Դիտարկվող մոտեցման մեջ նաև առաջարկվում է անցկացնել նախնական տվյալների վերլուծություն (դիսկրետացում, վարկառուի վիճակագրորեն նշանակալի բնութագրերի որոնում) և օգտագործել որակի ընտրության տարբեր չափանիշներ՝ օպտիմալ կառուցվածքը ստանալու համար: Ստացված արդյունքների հիման վրա բանկի հաճախորդները վարկունակության մակարդակով բաժանվում են մի շարք դասերի: Մշակված ալգորիթմի հիման վրա ստացվել է վարկունակության կանխատեսման արդյունավետ մեթոդ, որը թույլ է տալիս գնահատել վարկի վճարման հավանականությունը՝ վարկառուի հայտնի բնութագրերի հիման վրա:

Այսպիսով, մեքենայական ուսուցման տարբեր մոդելների կիրառումը թույլ է տալիս բարձրացնել վարկունակության կանխատեսման արդյունավետությունը, վարկային ռիսկի գնահատման որակը և օպտիմալացնել վարկի տրամադրման գործընթացը:

Հոդվածը տպագրության է երաշխավորել խմբագրական խորհրդի անդամ, տ.գ.դ. Թ.Ն. Մանասեյանը: