
ДИСКРИМИНАЦИОННА СПОСОБНОСТ ПРИ ОЦЕНКАТА ЗА ВЕРОЯТНОСТ ОТ НЕПЛАТЕЖОСПОСОБНОСТ ПОСРЕДСТВОМ НЯКОИ АЛГОРИТМИ ЗА МАШИННО САМООБУЧЕНИЕ

Антонио В. Дичев¹

¹Стопанска академия „Д. А. Ценов“ – Свищов, България

E-mail: ¹antoniodichev@yahoo.com

Резюме: Изследването прави опит да подчертае значението и добавената стойност на някои алгоритми за машинно самообучение при оценяването на вероятността от неплатежоспособност. Резултатите от изследването акцентират върху допълнителната дискриминационна способност, но са налице съществени и множество други аспекти от прилагането на машинното самообучение за оценка на кредитния риск. Тези аспекти могат да бъдат обособени като конкретни възможности и предизвикателства. По отношение на дискриминационната способност при използваната извадка резултатите свидетелстват за превъзходството на машинното самообучение спрямо традиционно установените и познати модели. За отделните бизнес организации с експозиция към кредитния риск това би могло да допринесе до по-малко кредитни загуби при по-голям обем от бизнес операции.

Ключови думи: вероятност от неплатежоспособност, машинно самообучение, риск оценка, кредитен риск.

Тази статия се **цитира**, както следва: **Дичев, А. (2023).** Дискриминационна способност при оценката за вероятност от неплатежоспособност посредством някои алгоритми за машинно самообучение. Народностопански архив, (4), с. 17-30.

URL: nsarhiv.uni-svishtov.bg

DOI: <https://doi.org/10.58861/tae.ea-nsa.2023.4.02.bg>

JEL: B23, C58, G32.

* * *

Добре изградената рамка за управление и оценка на риск е доказала в исторически план своето ключово място във финансите. Използването на прецизни модели, които отразяват максимален брой фактори и информация, е ключов елемент в постигането на обективна рамка за управление на риска с минимални елементи на субективност. При анализа на големи масиви от данни откриването на по-трудно забележими зависимости и създаване на риск модели с отличаваща се висока прецизност съществена роля в дигиталното съвремие има т.нар. ”машинно самообучение¹”(machine learning).

В тази връзка **обект** на изследване в настоящата работа е машинното самообучение на данни. **Предмет** на изследването е оценяването на вероятността от неплатежеспособност² (PD). **Изследователската теза** се определя, както следва: Посредством прилагането на методи за машинно самообучение при оценяването на вероятността от неплатежеспособност може да се постигне допълнителна прецизност.

Целта на изследването е свързана с апробирането на машинното самообучение като инструмент с добавена стойност при управлението на кредитния риск. Основна **задача** на разработката е извеждане на ефекта на някои алгоритми за машинно самообучение върху дискриминационната способност при оценките за вероятност от неплатежеспособност.

1. Теоретични основи и преглед на литературните източници по темата

1.1. Общо разбиране за машинно самообучение (ML)

В множеството достъпни ресурси често се използва следната дефиниция за машинно самообучение (ML):

„Една компютърна програма се учи от опит E по отношение на определен клас от задачи T и измерител за коректност P , ако при изпълнението на задачи T , оценено посредством P , се наблюдава подобрене при натрупването на исторически опит.” (Mitchell, 1997, p. 2).

¹ Заб.: В текста да се разбират за еднозначни термините машинно самообучение - machine learning (ML)

² Заб.: В текста да се разбират за еднозначни термините: вероятността от неплатежеспособност - probability of default (PD).

Въпреки че логистичната регресия отговаря на дефиницията по-горе и в теоретичен аспект попада под обхвата на машинното самообучение, изследването разглежда логистичната регресия като алтернатива на алгоритмите за машинно самообучение. Тоест, тя се разглежда като представител на вече наложилите се традиционни алгоритми и методи за моделиране на вероятността от неплатежоспособност (PD). Това разграничение се прави, тъй като в практиката често логистичната регресия не бива припознавана и разглеждана като ML алгоритъм. Това разделение на автора е предложено и от актуалните регулаторни и институционални дискусии в банковия сектор (EBA, 2021).

1.2. Вероятност от неплатежоспособност и нейното място в кредитния риск

Според Базелския комитет по банков надзор (BCBS, 2000) кредитният риск е рискът от загуба, произтичаща от неизпълнение на договорните задължения на кредитополучателя, което води до финансова загуба за кредитора. Той включва както потенциала за неизпълнение на задълженията, така и потенциала за влошаване на кредитното качество. Saunders, A., Cornett, M. (Saunders & Cornett, 2015, p. 288) разбират риска от неизпълнение като потенциална загуба в резултат на неспособността или нежеланието на кредитополучателя да изплати дългово задължение, както е договорено.

Тези дефиниции наблягат на потенциала за финансови загуби поради неуспех на кредитополучателите да изпълнят своите задължения за изплащане или влошаване на тяхната кредитоспособност. Обичайно и при неговото управление, кредитният риск разглежда както потенциала за пълно неизпълнение, така и за частична загуба поради намаляване на кредитоспособността на кредитополучателя.

Като част от рамката за управление на кредитния риск, финансовите институции, например, често използват статистически модели и методи за оценка на кредитния риск. Най-значимите показатели, които се оценяват посредством тези подходи, са PD (Probability of default), LGD (Loss given default) и EAD (Exposure at default).

- Probability of default/Вероятност от неплатежоспособност (PD) – един от ключовите показатели в областта на кредитния риск. Представя риска определен контрагент (физическо или юридическо лице) да изпадне в неплатежоспособност за определен период от време. В банковия сектор, например, показателят се използва както за вътрешни цели (вземане на решение за поемане на кредитен риск, ценообразуване и др.), така и при определянето на необходимия капитал за регулаторни цели.

- Loss given default/Загуба при неплатежоспособност (LGD) – освен вероятността да изпадне в неплатежоспособност, за бизнес организациите е важно да измерят и загубата при реализирането на този риск (изпадане в неплатежоспособност). LGD изразява точно това – % загуба в случай на неплатежоспособност. Показателят е повлиян от множество фактори, както индивидуални на конкретния контрагент (бизнес модел, капиталова структура и др.), така и специфични за сделката (обезпечения, условия и т.н.).

- Exposure at default / Експозиция при неплатежоспособност (EAD) – изразява очакваната експозиция на конкретния контрагент към момента на настъпване на кредитно събитие (неплатежоспособност). Сумата представлява максималната стойност на потенциалната загуба от съответния контрагент.

1.3. Преглед на изследванията за машинно самообучение като метод за оценка на кредитния риск

В популярната литература, свързана с темата за машинното самообучение, съществуват редица проучвания. Сред по-популярните изследвания, оценяващи кредитния риск, попада разработката на Milojevic, N., Redzepagic, S. (Milojevic & Redzepagic, 2020). Те изследват изкуствения интелект и машинното самообучение като възможно решение за подобрене на риск мениджмънта в банковия сектор. Авторите заключават, че е налице потенциал за положителен ефект от използването на ML, особено в следните области от риск мениджмънта: кредитен риск, пазарен риск, ликвиден и операционен риск. Изследването подчертава и редица предизвикателства и отворени въпроси. Обръща се внимание на възможността за бъдещи академични разработки по темата за ML и неговото приложение при управлението на риска. Gunasilan, U., Sharma, R. (Gunasilan & Sharma, 2022) констатират, че машинното обучение в управлението на риска в сектора на финансовите услуги все още е недостатъчно проучено. Въпреки че има много проучвания за кредитните рискове, на други рискове като ликвидни рискове, пазарни рискове и операционни рискове е обърнато минимално внимание. Трудът установява, че приложенията за машинно обучение имат потенциала да разработят по-ефективни модели за управление на риска. Освен това емпирично се доказва, че техниките за машинно обучение в управлението на риска осигуряват по-добри и поточни резултати от традиционните статистически модели. Очертават се и някои области, които се нуждаят от допълнително проучване. Например документът предлага по-задълбочени последващи проучвания на модели за машинно обучение при различните видове банкови рискове. Dowling,

M., Aziz, S. (Dowling & Aziz, 2019) изследват как решенията в областта на изкуствения интелект (AI) и машинното обучение (ML) (Zarkova, Kostov, Angelov, Pavlov, & Zahariev, 2023) трансформират управлението на риска. Заклучават с оптимистична картина за ролята на изкуствения интелект и машинното обучение в управлението на риска, но се отбелязват някои практически ограничения, особено около подходящите политики за мениджмънт на данни, както и липсата на необходимия набор от умения, налични в корпоративния свят. Huang, B., Wei, J., Tang, Y., Liu, C. (Huang, Wei, Tang, & Liu, 2021) прилагат няколко ML метода за оценка на корпоративния риск. В понятието корпоративен риск авторите разглеждат само кредитния риск. Използваните алгоритми от авторския колектив са методите на: случайната гора (Random forest (RF)) (Prodanov, Angelov, & Zarkova, 2022), опорните вектори (Support vector machines (SVM)) (Zahariev, et al., 2020a), (Zahariev, Angelov, & Zarkova, 2022a) и адаптивно подсилване (AdaBoost). Заклучението на изследването е, че и трите приложени алгоритми за машинно обучение могат ефективно да оценят рисковете на предприятието. Shi, S., Tse, R., Luo, W., D'Addiba, S., Pau, G. (Shi, 2022) очертават фундаменталната роля на кредитния риск и правилната му оценка при съвременните икономики. Авторският състав предлага нова методология за класификация на управлявани от машинното обучение алгоритми за кредитен риск. Резултатите от разработката показват, че: 1) повечето модели за дълбоко обучение превъзхождат традиционните статистически алгоритми при оценката на кредитния риск и 2) ансамблови методи („ensemble learning“) осигуряват по-висока точност в сравнение с единичните модели.

Настоящият преглед на литературните източници не изчерпва наличната литература по темата, а предоставя теоретична база за последващи теоретико-емпирични изследвания.

2. Описание на данните и използваната методология в изследването

2.1. Извадка от данни и ключови обработки

Използваната извадка (Home Credit Group, n.d.) от данни представлява публичната база данни на финансова институция „Home Credit“. Всяко наблюдение съдържа информация за кредитоискателя, желанието от него кредит и променлива, съдържаща два класа за съответния клиент (кредитоискател). Тази бинарна променлива показва дали впоследствие кредитоискателят е изпаднал в неплатежоспособност (default). Извадката

е сравнително популярна поради своята достъпност, голямо количество от наблюдения и индикатори. Съдържат се анонимни данни с показатели като: възраст, пол, доход, вноски, размер на кредита, трудов стаж, притежание на имущество и други. Общият брой наблюдения възлиза на 307 511 бр. На база първоначално наличните данни са конструирани допълнителни индикатори като % на усвояемост по кредитна карта, сума на тегления, сума на плащания и други. Финалният брой на характеристиките за клиента и апликацията (независими променливи) възлиза на 288 бр. В публикуваните данни се съдържат няколко отделни файла с данни. Използваните файлове от автора са: “application_train.csv”, “credit_card_balance.csv”, “POS_CASH_balance.csv” и “installments_payments.csv”. Запазени са наблюденията, при които за една или повече от характеристиките липсва информация. Присвоена е специална категория при наличие на липсваща стойност. Важно е да се отбележи, че съществува разнообразие от методи за тяхното прогнозиране или заместване. Добавената стойност от подобен тип подход, както и най-подходящият метод, би могло да се разгледат в отделна разработка. Извадката е допълнително разделена на т.нар. разработваща (70%) и тестова (30%) извадка от данни. Разработващата е използвана в процеса на моделиране, а валидационната за оценка и потвърждаване на дискриминационна способност върху данни, които не са участвали в процеса по изграждане на модела. Разделението е на случаен принцип, запазвайки съотношението между класовете на зависимата променлива.

2.2. Използвани алгоритми и методи за моделиране на PD

В практиката вероятността от неплатежеспособност се разглежда като бинарен класификационен проблем. Тоест, налични са две групи клиенти и целта на модела е да ги разграничи една от друга, с възможно най-висока точност. В контекста на оценката за кредитен риск, в частност риск от неплатежеспособност, разделението на двете групи се дефинира, както следва:

- клас 0 – клиенти, при които **не е настъпвало** събитие на неплатежеспособност за определения времеви хоризонт (обичайно 1 година);
- клас 1 – клиенти, при които **е настъпвало** събитие на неплатежеспособност за определения времеви хоризонт (обичайно 1 година).

В банковия сектор инструментите за оценка на риска от неплатежеспособност за физически лица често са наричани скоринг модели, а при сегмента на юридическите лица – рейтинг модели. Налични са и други характерни различия между двете категории модели, но предметът на

настоящото изследване не предполага техния по-задълбочен преглед и синтез.

Използваните алгоритми за моделиране на PD като оценка за кредитен риск са логистична регресия (Logistic regression), класификационно и регресионно дърво (Classification And Regression Tree), метод на екстремното градиентно подсилване (Extreme Gradient Boosting) и методът на градиентно подсилване (Gradient Boosting). Обектът и обхватът на настоящото изследване не предполагат детайлен преглед на допусканията и спецификите при различните ML алгоритми, вкл. логистичната регресия. Следва да се подчертае, че и двете групи модели (традиционни модели и алгоритми за машинно самообучение) не изчерпват подходящите и текущо прилагани методи за предиктивни класификационни проблеми.

2.2.1. Традиционни алгоритми и методи

- логистична регресия (Logistic regression) – инструментът е много добре познат в теорията и традиционно използван метод в практиката. В настоящото изследване се прилага логистична регресия от биномен вид, тоест зависимата променлива има два класа. Линеиният израз на регресионния алгоритъм е, както следва (Park, 2013):

$$\text{logit}(y) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_k X_k \quad (1)$$

- WoE трансформация и стандартизация

Изграждането на модел чрез логистична регресия (1) е извършено посредством подхода на групиране при независимите променливи и т.нар. WoE (Weight of evidence) трансформация, вкл. изчисляване на индикатора IV (information value) (Siddiqi, 2006, pp. 77-81). Този подход е много широко прилаган и утвърден в практиката. Сред основните предимства на подхода са лесна интерпретация, проследяване за монотонна и логична връзка между трансформираните независими променливи и кредитния риск. Подходът спомага и при избора на финалната архитектура от независими променливи в регресионното уравнение и други.

2.2.2. Machine learning алгоритми

- стандартизация – приложен е т.нар. метод на ”стандартизация” на отделните характеристики от извадката. Методът е честа практика в

машинното самообучение³. Методът на стандартизация е широко използван и прилаган в практиката и литературата, а използваният в настоящото изследване метод е т.нар. ”z-score” (Aksu & Güzeller, 2019):

$$z_{i,j} = (x_i - u_i) / s_i \quad (2)$$

$z_{i,j}$ = стойност на наблюдение j , променлива i ;

u_i = средна стойност, променлива i ;

s_i = стандартно отклонение, променлива i .

- хиперпараметри – в машинното самообучение има т. нар. хиперпараметри. Това са настройки на конкретния алгоритъм, които не се “научават” от самите данни и които определят поведението и колко сложна да бъде архитектурата на модела. Тоест, дори при използването на един и същи алгоритъм (напр. XGBoost) е съвсем нормално при различни извадки от данни да има различна архитектура и сложност на финалния модел. До финалните стойности на хиперпараметрите се достига посредством процес по търсене на техните най-оптимални такива. Това е съществена и ключова стъпка от цялостния процес по прилагане на ML алгоритми. В най-общ смисъл правилото е да се избере тази комбинация от стойности на настройките, която води до най-добри финални резултати. Съществуват различни подходи за търсене на оптималните стойности (grid search, random search и други). В настоящата разработка е използван т. нар. random search подход.

- Алгоритъм на класификационно и регресионно дърво (Classification And Regression Tree, CART) – термин, използван от (Leo & Friedman, 1984), който на практика представлява т. нар. “дърво на решенията” (decision tree). Инструментът представя различните последиствия (резултати) от различни сценарии и варианти на развитие. В контекста на моделирането тези сценарии се дефинират посредством разделяне на извадката от данни на база стойностите на показателите вътре в нея. Нивата и броят на деления, базирани върху данните се извършват посредством минимизирането на разликата между резултатите от ”дървото на решения” и реалните данни.

CART е добре познат в литературата и практиката и е основа, върху която са изградени едни от най-модерните и важни ML алгоритми, например: RF, XGBoost, GB и други.

- алгоритъм на градиентното подсилване (Gradient Boosting, GB) – в резюме, алгоритъмът наподобява изключително много XGBoost и

³ Виж повече (Aksu & Güzeller, 2019).

също принадлежи към семейството на "ensemble learning" алгоритмите. XGBoost е оптимизиран откъм бързодействие на процеса по оптимизация на хиперпараметрите, които в по-голямата си част са едни и същи при двата алгоритъма. Освен разликата в ефективността на оптимизация при хиперпараметрите, GBC няма вградени методи за т. нар. регуляризация⁴.

- алгоритъм на екстремното градиентно подсилване (Extreme Gradient Boosting, XGBoost) (Chen & Guestrin, 2016) – един от най-мощните и широко използвани ML алгоритми днес. Методът принадлежи към семейството на т.нар. "ensemble learning" алгоритми. Тази група алгоритми се нарича така, защото в архитектурата си комбинира множество по-слаби индивидуални модели (често това е алгоритъмът "дърво на решенията"). Посредством итеративно добавяне на по-слаби модели, от своя страна, се нарича "gradient boosting". Популярността на алгоритъма се дължи основно на неговата ефективност и ефикасност в най-различни сфери от реалния свят. Алгоритъмът може да бъде прилаган както за задачи, изискващи регресия, така и такива от класификационен характер, какъвто е PD оценката.

2.3. Оценка на дискриминационната способност (Somers'D/AUC)

В статистическия анализ тази характеристика се използва, за да измери доколко моделът успява да разграничи различните групи клиенти посредством своята оценка. Скоринг моделите са подходящ пример, защото същите се опитват да ранкират коректно отделните клиенти спрямо техния риск от неплатежоспособност. Идеалният модел би разпределил така клиентите, че да е възможно посредством оценката от модела да се отделят всички клиенти, които изпадат в неплатежоспособност от останалите.

Важно е да се отбележи, че в реалния свят и практиката наблюдаваните данни са "небалансиранни" (единият от двата класа е по-често срещан в извадката или популацията). От своя страна това налага използването на подходяща метрика, която да оцени правилно коректността на изградения инструмент за PD оценка.

Somers' D е асиметрична мярка за връзката между две променливи, която играе централна роля като параметър за ранг или „непараметричните“ статистически методи. При независима променлива X и зависима Y , може да се калкулира оценка $d_{Y/X}$, която представлява мярка за ефекта на

⁴ Методи за избягване на т. нар. overfitting. Виж повече за L1 (Lasso) и L2 (Ridge) подходи за регуляризация.

X върху Y . Тоест, $d_{Y/X}$ може да се използва като индикатор, доколко успешно X предсказва Y . Калкулацията му се извършва посредством сортиране на двете променливи (X и Y) и съгласно следния израз (Göktaş & Öznur, 2011, p. 26):

$$d_{Y/X} = \frac{P-Q}{D_r} \quad (3)$$

$P = \sum_{i,j} f_{ij}C_{ij}$ = Вероятността случайно избрана двойка наблюдения да имат еднозначна подредба.

$Q = \sum_{i,j} f_{ij}D_{ij}$ = Вероятността случайно избрана двойка наблюдения да имат противоположна подредба.

$C_{ij} = \sum_{k>i} \sum_{l>j} f_{kl} + \sum_{k<i} \sum_{l<j} f_{kl}$ = Общ брой съгласувани (concordant) двойки.

$D_{ij} = \sum_{k>i} \sum_{l<j} f_{kl} + \sum_{k<i} \sum_{l>j} f_{kl}$ = Общ брой несъгласувани (discordant) двойки.

f_{ij} = общ брой наблюдения при ред i и колона j от таблица

$$D_r = W^2 - \sum_{j=1}^C r_j^2$$

r_j = общ брой от честотното разпределение за ред j на двойно сортираната cross таблица

$$W = \sum_{j=1}^C c_j = \sum_{i=1}^R r_i = \text{общо наблюдения в извадката}$$

Стойността на Somers' D е в интервала от -1 до 1. Стойност от -1 означава 100% негативна връзка, а 1 означава 100% положителна връзка между независима и зависима променлива. Стойност от 0 означава, че липсва връзка (Göktaş & Öznur, 2011, p. 26).

Настоящата разработка ще използва именно Somers' D коефициента като основна метрика за дискриминационната способност. Често в литературата и практиката се използва и т.нар. AUC (Area under the curve), чиято връзка със Somers' D е, както следва (ЕСВ, 2019, p. 68):

$$AUC = (d_{Y/X} + 1)/2 \quad (4)$$

При различните типове модели изискваното и очаквано минимално ниво на дискриминационна способност е различно. Предвид това, че в настоящата разработка акцентът ще бъде основно върху това, дали и с колко ML методите допринасят за по-висока дискриминационна способност, ще разглеждаме изменението на показателя спрямо традиционните алгоритми (логистична регресия).

3. Обобщение и анализ на резултатите

В резултат на извършеното апробиране Таблица 1 представя сравнителна таблица на оценката за дискриминационна способност между всички приложени алгоритми:

Таблица 1.

Дискриминационна способност (тестова извадка)

Алгоритъм	AUC	Somers'D
Logistic regression	0.66	0.33
CART	0.67	0.33
GB	0.74	0.49
XGBoost	0.77	0.54

Източник: Изчисления на автора.

Резултатите показват, че при тази извадка приложените ML алгоритми превъзхождат логистичната регресия по отношение на дискриминационната способност. Добавената стойност в дискриминационната способност е съществена, особено при алгоритъма с най-добри резултати (+ 0.21 Somers'D спрямо логистичната регресия).

За да се изрази добавената стойност през призмата на финансовите резултати за една компания, се проследява изменението в т. нар. грешки от първи род (type I errors/false positives) и грешки от втори род (type II errors/false negatives). В контекста на настоящото изследване грешките от първи род изразяват броя наблюдения (клиенти), които са в редовност, но "маркирани" от модела като такива в нередовност клиенти. От своя страна грешките от втори род изразяват броя клиенти в неизпълнение, които моделът идентифицира като редовни. В практиката е възможно да се придава различна тежест и значимост на двата типа грешки в зависимост от бизнес спецификите. При анализа на резултатите е присвоено еднакво тегло. Данните от тестовата извадка са налични в Таблица 2.

От резултатите в Таблица 2, прави впечатление, че XGBoost допринася в подобрение и при двата типа грешки. Тоест, спрямо традиционните алгоритми, прилагането на този алгоритъм ще позволи на една бизнес организацията да:

- наблюдава по-малко кредитни загуби (по-нисък брой false negatives);
- реализира по-голям обем бизнес взаимоотношения (по-нисък брой false positives).

Таблица 2.

Грешки от първи и втори род (тестова извадка)⁵

Алгоритъм	false positives	false negatives
Logistic regression	17 425	4 314
CART	11 871	4 851
GB	10 125	4 327
XGBoost	9 613	4 230

Източник: Изчисления на автора.

Подчертава се, че изведените резултати и изводи са валидни само за използваната извадка от данни и е възможно да не бъдат потвърдени върху други извадки.

Заклучение

В заключение на така извършеното проучване изследването подчертава значението и добавената стойност на ML техниките при оценката на вероятността от неплатежоспособност. По отношение на дискриминационната способност резултатите свидетелстват за превъзходство на тези решения спрямо традиционно установените. На база анализ данните от т.нар. матрица на грешките (confusion matrix) се идентифицира потенциал за подобрене в коректното разпознаване и при двата класа. За отделните бизнес организации с експозиция към кредитен риск това, от своя страна, би могло да допринесе до по-малко кредитни загуби, при по-голям обем от бизнес операции.

Резултатите и изследването акцентират върху допълнителната дискриминационна способност, но са налице съществени и множество други аспекти от прилагането на ML за оценка на кредитния риск. Тези аспекти могат да бъдат групирани като възможности и предизвикателства. Тъй като настоящата статия се фокусира само върху дискриминационната способност, може да се заяви, че са налични множество възможности и потребности за допълнителни изследвания за ролята, значимостта и предизвикателства пред ML като инструмент за оценка на риска. Акцентира се върху това, че резултатите се базират на конкретно използваната извадка и би могло да се извърши допълнително изследване относно ефекта от използване на алгоритми за машинно самообучение при различни извадки от данни.

⁵ Поместеният брой на false positives / false negatives са извлечени от точката на ROC кривата с най-добро съотношение между % на false positives / false negatives, за съответния модел.

ИЗПОЛЗВАНИ ИЗТОЧНИЦИ

- Aksu, G., & Güzeller, C. (2019). *The Effect of the Normalization Method Used in Different Sample Sizes on*.
- BCBS. (2000). *Принципи за управление на кредитния риск*.
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*.
- Dowling, M., & Aziz, S. (2019). *AI and machine learning for risk management*.
- EBA. (2021). *Discussion paper on machine learning for IRB models*.
- ECB. (2019). *Instructions for reporting the validation results of internal models*.
- Göktaş, A., & Öznur, İ. (2011). *A Comparison of the Most Commonly Used Measures of Association for Doubly Ordered Square Contingency Tables via Simulation*.
- Gunasilan, U., & Sharma, R. (2022). Risk management in banking using Machine learning techniques as an alternative analysis tool – a review. *International Journal of Management (IJM)*, 13(5), 73-78. doi:<https://doi.org/10.17605/OSF.IO/CTKUG>
- Home Credit Group. (15 09 2023 r.). *Home Credit Default Risk*. <https://www.kaggle.com/competitions/home-credit-default-risk/data>
- Huang, B., Wei, J., Tang, Y., & Liu, C. (2021). *Enterprise Risk Assessment Based on Machine Learning*.
- Leo, B., & Friedman, J. H. (1984). *Classification and Regression Trees*.
- Milojevic, N., & Redzepagic, S. (2020). *Prospects of Artificial Intelligence and Machine Learning Application in Banking Risk Management*.
- Mitchell, T. (1997). *Machine Learning*.
- Park, H. (2013). *An Introduction to Logistic Regression: From Basic Concepts*.
- Prodanov, S., Angelov, P., & Zarkova, S. (2022). Real estate in Bulgaria from the global financial crisis to the COVID-19 crisis. Effects of macro-prudential policy: evidence from Bulgaria. *87th International Scientific Conference of Economic and Social Development: Economics, Management, Finance and Banking* (pp. 162-170). Cakovec, Croatia: Economic and Social Development (Book of Proceedings), vol. 87.
- Saunders, A., & Cornett, M. M. (2014). *Управление на финансови институции: Подход за управление на риска*.
- Shi, S., Tse, R., Luo, W., D'Addiba, S., & Pau, G. (2022). *Machine learning-driven credit risk: a systemic review*.
- Siddiqi, N. (2006). *Credit Risk Scorecards. Developing and Implementing Intelligent Credit Scoring*.
- Zahariev, A., Angelov, P., & Zarkova, S. (2022a). Estimation of Bank Profitability Using Vector Error Correction Model and Support Vector

Regression. *Economic Alternatives*, 157-170.

doi:<https://doi.org/10.37075/EA.2022.2.01>

Zahariev, A., Zveryakov, M., Prodanov, S., Zaharieva, G., Angelov, P., Zarkova, S., & Petrova, M. (2020a). Debt management evaluation through support vector machines: on the example of Italy and Greece. *Entrepreneurship and Sustainability Issues*, 7(3), 2382-2393. doi:[https://doi.org/10.9770/jesi.2020.7.3\(61\)](https://doi.org/10.9770/jesi.2020.7.3(61))

Zarkova, S., Kostov, D., Angelov, P., Pavlov, T., & Zahariev, A. (2023, April 3). Machine Learning Algorithm for Mid-Term Projection of the EU Member States' Indebtedness. *Risks* 11(4). doi:<https://doi.org/10.3390/risks11040071>

Антонио В. Дичев е докторант в катедра „Финанси и кредит“, Стопанска академия „Д. А. Ценов“ – Свищов, България. **Научни интереси:** машинно самообучение, оценка на финансовия риск, оценка на кредитен риск.

ORCID ID: 0009-0005-2552-1819

ISSN 0323-9004

Народностопански архив

Свищов, година LXXVI книга 4 - 2023

**За обществената реалност, теорията
и обучението по икономика**

**Дискриминационна способност при оценката
за вероятност от неплатежоспособност
посредством някои алгоритми за машинно
самообучение**

**Възможности за преодоляване на проблемите
във фазата на изплащане на пенсиите при
старост от универсален пенсионен фонд**

**Интелектуалният капитал и пазарната
капитализация на публичните дружества**

**Финанси и Стопанска отчетност – факултети,
научни изследвания и 70-годишна традиция**

СТОПАНСКА АКАДЕМИЯ „Д. А. ЦЕНОВ“



СВИЩОВ

РЕДАКЦИОНЕН СЪВЕТ:

Проф. д-р Андрей Захариев – главен редактор
Проф. д-р Йордан Василев – зам. главен редактор
Проф. д-р Стоян Проданов
Проф. д-р Тодор Кръстевич
Доц. д-р Искра Пантелеева
Доц. д-р Пламен Йорданов
Доц. д-р Светослав Илийчовски
Доц. д-р Пламен Петков
Доц. д-р Анатолий Асенов

МЕЖДУНАРОДЕН СЪВЕТ:

Проф. д-р ик.н. Михаил А. Ескиндаров – Финансов университет при Правителството на Руската федерация, Москва (Русия).
Проф. д-р ик.н. Григоре Белостечник – Молдовска академия за икономически изследвания, Кишинев (Молдова).
Проф. д-р ик.н. Михаил Ив. Зверяков – Одески държавен икономически университет, Одеса (Украйна).
Проф. д-р ик.н. Андрей Крисоватий – Тернополски национален икономически университет, Тернопол (Украйна).
Проф. д-р ик.н. Йон Кукуй – Университет Валахия, Търговище (Румъния)
Проф. д-р Кен О'Нийл – Университет Ълстер (Ирландия)
Проф. д-р Ричард Торп – Университет Лийдс (Великобритания)
Проф. д-р ик.н. Олена Непочатенко – Умански национален аграрен университет, Уман (Украйна)
Проф. д-р ик.н. Дмитрий Лукьяненко – Киевски национален икономически университет „Вадим Гетман”, Киев (Украйна)
Доц. д-р Мария Стефан – Университет „Валахия”, Търговище (Румъния)
Доц. д-р Анисоара Дунка – Университет "Валахия", Търговище (Румъния)
Доц. д-р Владимир Климуk – Брановички държавен университет, Бранович (Беларус)

Екип за техническо обслужване:

Технически секретар: д-р Росица Проданова
Стилов редактор: Анка Танева
Превод английски: ст. преп. Венцислав Диков и ст. преп. д-р Маргарита Михайлова

Адрес на редакцията:

5250 Свищов, ул. „Ем. Чакъров” 2
Проф. д-р Андрей Захариев – главен редактор
☎ (+359) 889 882 298
Д-р Росица Проданова – технически секретар
☎ (+359) 631 66 309, e-mail: nsarhiv@uni-svishtov.bg
Благовеста Борисова – компютърен дизайн
☎ (+359) 882 552 516, e-mail: b.borisova@uni-svishtov.bg

Отпечатването на списанието за 2023 г. се осъществява с безвъзмездната финансова помощ на Фонд „Научни изследвания” – Договор ДНП № КП-06-НП4-36 по конкурс „Българска научна периодика – 2023 г.”

© Академично издателство „Ценов” – Свищов
© Стопанска академия „Димитър А. Ценов” – Свищов

НАРОДНОСТОПАНСКИ АРХИВ

ГОДИНА LXXVI, КНИГА 4 – 2023

СЪДЪРЖАНИЕ

Методи Кънев

За обществената реалност, теорията и обучението по икономика /3

Антонио В. Дичев

Дискриминационна способност при оценката за вероятност от неплатежоспособност посредством някои алгоритми за машинно самообучение /17

Йордан С. Йорданов

Възможности за преодоляване на проблемите във фазата на изплащане на пенсиите при старост от универсален пенсионен фонд /31

Венцислав Диков

Интелектуалният капитал и пазарната капитализация на публичните дружества /51

Андрей Захариев, Атанас Атанасов, Галина Чиприянова

Финанси и Стопанска отчетност – факултети, научни изследвания и 70-годишна традиция /73