International Scientific Conference 20 UNITECH 26

INTERNATIONAL SCIENTIFIC CONFERENCE
19-21 November 2026
GABROVO

КАТЕГОРИЗАЦИЯ НА ИНВЕСТИЦИОНЕН РИСК ЧРЕЗ ДИСКРИМИНАНТЕН АНАЛИЗ, K-NN, SVM, FFNN, CFNN И PNN ПОДХОДИ

Ивелина Балабанова[1], Мария-Елевтерия Филиаду[2]
[1]Технически университет - Габрово, ул. „Хаджи Димитър“ № 4, Габрово, България
[2]Технически университет - Габрово, ул. „Хаджи Димитър“ № 4, Габрово, България
ivstoeva@abv.bg

INVESTMENT RISK CATEGORIZATION THROUGH DISCRIMINANT ANALYSIS, K-NN, SVM, FFNN, CFNN AND PNN TECHNIQUES

Ivelina Balabanova[1], Maria-Eleftheria Filiadou[2]
[1]Technical University of Gabrovo, 4 "Hadji Dimitar" Str., Gabrovo, Bulgaria
[2]Technical University of Gabrovo, 4 "Hadji Dimitar" Str., Gabrovo, Bulgaria
ivstoeva@abv.bg

Abstract
The paper presents a methodology design processes for investment risk categorization regarding to the volume of funds increasing for digitization of integrated and future Education ICT infrastructures in University organizations. The methodology have been developed on the basis of Machine Learning and Artificial Neural Networks. Linear and Quadratic Discriminant classifiers with the highest efficiency of 94.165 % were evaluated. Classification models with Euclidean, Euclidean Squared, Cityblock and Chebychev metric distance on the basis of k-Nearest Neighbors method have been analyzed with the highest accuracy obtained, respectively 84.444 %. Nu-based Support Vector Machine models with Linear, Polynomial, RBF and Sigmoid functions were examined with the highest efficiency 96.667 %. A comparative analysis about Feed-Forward, Cascade-Forward and Probabilistic Neural Networks with the best network performance 98.300% have been synthesized.

Keywords: Risk Categorization; Discriminant Analysis; k-NN; Nu-SVM; Feed-Forward; Cascade-Forward; Probabilistic Neural Networks.

1. ВЪВЕДЕНИЕ

От една страна пазарните условия са в състояние да формират нестабилна среда за появата на негативни тенденции при оперирane с финансови инструменти и загубата на активи и потенциални инвеститори [1, 2]. Като обект на икономически анализ инвестиционните рискове обхващат следните компоненти, съответно "Market", "Inflasion", "Credit, "Liquidity", "Interest Rate". От друга, един от ключовите критерии за изграждане на висококвалифицирани специалисти и маркер за напредъка на едно общество е интегрирането на образователни инвестиции на нивото на международно образователно пространство [3-4]. Ето защо ключов критерии за успех е установяването на симбиоза между:

- ❖ насърчаването на инвестиционни образователни програми на държавническо ниво, свързани с иновациите в информационно-комуникационните технологии;
- ❖ адекватна оценка на факторите на бизнес средата и състоянието на местната икономика;
- ❖ интегрирането на новите цифрови технологии и адекватни регулатор-

ни рамки относно изразходването на средства [5-8].

2. КОМПОЗИЦИОННА ЗАДАЧА НА ИЗСЛЕДВАНИЯТА

Предметът на изследване в настоящия доклад е насочен към анализ на риска относно повишаване на инвестиционния праг на средствата за обновяване на информационните инфраструктури в образователните институции по региони. Финансирането е предвидено от страна на държавни агенции и министерства към областни и общински центрове. Анализът на риска е базиран на отчитане и оценка на следните инокомически фактори:

- ❖ Market Volatility Index (MVI);
- ❖ Inflation Rate (IR);
- ❖ Stock Index Change (SIC);
- ❖ Investment Duration Days (IDD),

обвързани с подобряването/влошаването на степента на интеграция на цифровите технологии за всеки регион по отношение на вече предоставени и изразходени бюджетни финансови средства. За целта е предложен подход, включващ методи и алгоритми на Машинното обучение в единна концепция за категоризация на два рискови прага, съответно:

- ❖ Средна степен на риск;
- ❖ Висока степен на риск.

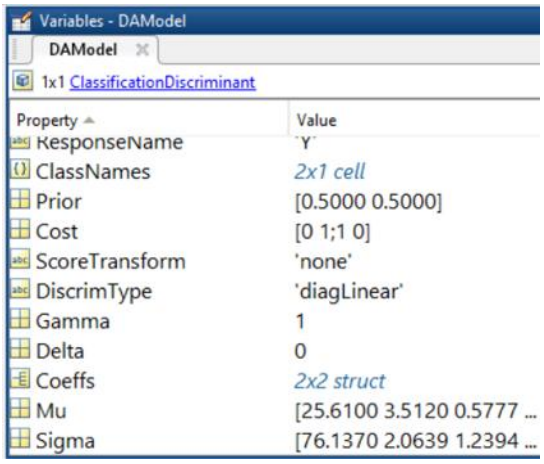
Разработеният подход предвижда използването на следните аналитични апарати: Discriminant Analysis (DA); k-Nearest Neighbors (k-NN) и Support Vector Machine (SVM); Feed-Forward, Cascade-Forward и Probabilistic Neural Networks (FFNNs, CFNNs и PNNs).

3. КАТЕГОРИЗАЦИЯ НА РИСК ЧРЕЗ ДИСКРИМИНАНТНИ МОДЕЛИ В MATLAB

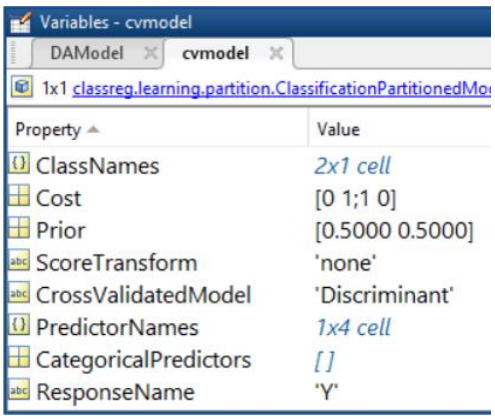
Проведено е изследване на създадени линейни и квадратични дискриминантни класификатори за категоризация на нива на риска, чиито резултати са обобщени в Таблица 1. Изследването разглежда приложимостта на шест типа дискриминантни модели, респективно: “Linear”, “Diagonal Linear”, “Pseudo-Linear”, “Quadratic”, “Diagonal Quadratic” и “Pseudo-Quadratic”. Направената оценка на ефективността е базирана на „Частта на некоректни класификации“ и калкулираната спрямо нейните нива „Класификационна точност“.

Таблица 1. *Дискриминанни класификатори за инвестиционен риск*

Тип клаисифи-катор	*Точност от Ресиб., %*	*Точност от Крос-валидиране, %*
Linear	91.67	90.00
Diagonal Linear	95.00	93.33
Pseudo- Linear	91.67	90.00
Quadratic	95.00	90.00
Diagonal Quadratic	93.33	93.33
Pseudo-Quadratic	95.00	90.00
Приблизително очаквана точност за “Diagonal Linear” модел: 94.165 %		



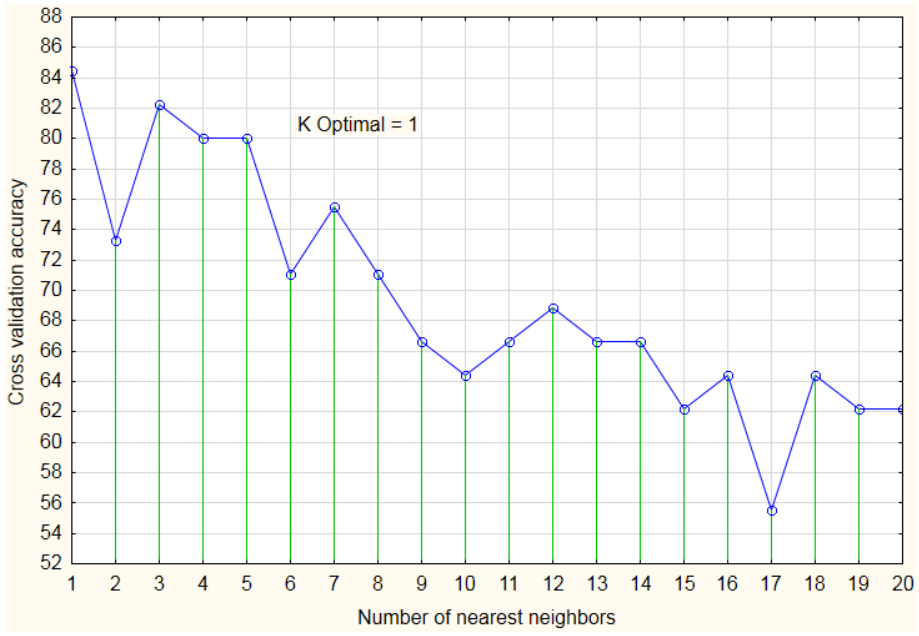
а)



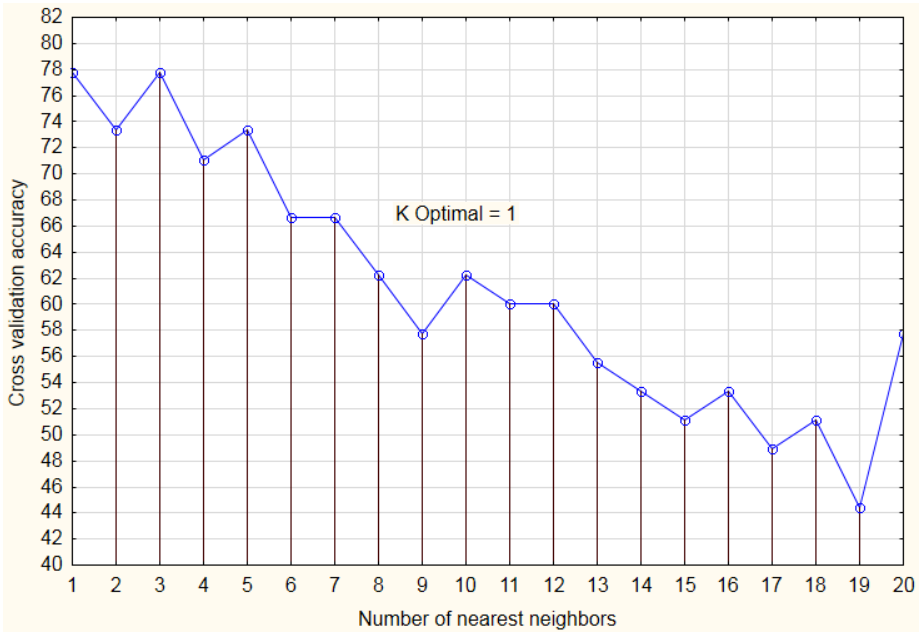
б)

Фиг. 1*. Синтез а) и валидиране б) на “Diagonal Linear” класификационен модел*

Въведени като подхода за оценка на качеството на класификация са „Ресубституция“ и „Крос-валидиране“. По отношение на всички разглеждани дискриминантни типове са получени точности над 90.00 %. Най-ниски прагове 91.67 % и 90.00 % са отчетени при “Linear”, “Pseudo-Linear” за Resubstitution и “Linear”, “Pseudo-Linear”, “Quadratic” и “Pseudo-Quadratic” Дискриминантни класификатори в случая на Cross-Validation подход. Според изложените резултати най-висока ефективност е достигната при вариант “Diagonal Linear”, отъждествявана с еквивалент 95.00 % при първия подход за оценка. Идентични най-добри класификационни точности 93.33 % са наблюдавани за “Diagonal Linear” и “Diagonal Quadratic” класификатори спря-



а)



б)

***Фиг. 2**. Класификационна точност от Крос-валидиране за k-NN модели при метрично разстояние а) “Euclidean / Euclidean squared”и б) “Chebychev”*

мо втория използван подход за оценка на качеството.

Във връзка с поставената класификационна задача е избран Diagonal Linear Discriminant тип класификатор, удовлетворяващ условие за гарантиране на "Approximately Expected Accuracy" при манипулации с нови данни 94.165 %. Списък на изходните метрики, потвърждаващ процедурите по създаване и валидиране на дискриминантния модел, е представен на Фиг. 1.

4. КАТЕГОРИЗАЦИЯ НА РИСК ПОСРЕДСТВОМ K-NN И SVM МЕТОДИ В СРЕДА

Създадени са k-NN класификационни модели на основата на "Euclidean", "Euclidean squared", "Cityblock" и "Chebychev" метрични единици за разстояние. Приложена е процедура по V-fold валидиране във връзка с оценка на качеството на моделите в границите на нарастване на "k-параметър" от 1 до 20. Резултатите от направената оценка са показани в графична форма на Фиг. 2.

Установено е еднакво поведение на критерия „Класификационна точност" при "Euclidean", "Euclidean squared" метрични разстояния. Изчислена е най-висок ранк 84.44444 % при определена оптимална стойност "k = 1". Минимално понижение на точността е регистрирано при метрично разстояние "Cityblock (Manhattan)", където е постигната максимална точност 82.22222 % при калкулирана оптимална стойност "k = 4". Спад на метриката под праг от 80.00 % е констатиран при k-NN модел с "Chebychev" тип разстояние. Тук е наблюдавана най-ниска точност в сравнение с останалите разглеждани класификатори 77.77778 %. Проведеният анализ дава основание за селекция на най-подходящ k-NN класификационен модел при задаване на "Euclidean" или "Euclidean squared" разстояния.

Анализирана е ефективността на ком-

```
Support Vector machine results:
  SVM type: Classification type 2 (nu=0,300)
  Kernel type: Polynomial (degree=3, gamma=0,250, coefficient=0,000)
  Number of support vectors = 19 (10 bounded)
  Cross-validation accuracy (%) = 95,556
  Support vectors per class: 11 (RiskLevelOne), 8 (RiskLevelTwo),

  Class. accuracy (%) = 95,556(Train), 93,333(Test), 95,000(Over-
```

а)

```
Support Vector machine results:
  SVM type: Classification type 2 (nu=0,200)
  Kernel type: Radial Basis Function (gamma=0,250)
  Number of support vectors = 13 (5 bounded)
  Cross-validation accuracy (%) = 91,111
  Support vectors per class: 6 (RiskLevelOne), 7 (RiskLevelTwo),

  Class. accuracy (%) = 97,778(Train), 93,333(Test), 96,667(Over-
```

б)

```
Support Vector machine results:
  SVM type: Classification type 2 (nu=0,300)
  Kernel type: Sigmoid (gamma=0,250, coefficient=0,000)
  Number of support vectors = 18 (10 bounded)
  Cross-validation accuracy (%) = 91,111
  Support vectors per class: 9 (RiskLevelOne), 9 (RiskLevelTwo),

  Class. accuracy (%) = 91,111(Train), 86,667(Test), 90,000(Over-
```

в)

***Фиг. 3**. Синтез на Nu-SVM класификатори при а) "Polynomial", б) "Radial Basis Function"и в) "Sigmoid" Кернел функции*

плекс от Nu-SVM класификатори, за които е калкулирана точност при обучение, тестови процесите и крайна метрика. Създадени са модели при дефиниране на различни типове Kernel функции, съответно “Linear”, “Polynomial”, “Radial Basis Function (RBF)” и “Sigmoid”, показани на Фиг. 3. Изчислени са базисните support vectors за всеки от посочените класификатори, респективно в последователен ред на изложените функционални типове „17“, „19“, „13“ и „18“.

Отчетени са следните класификационни метрики:

- Class. Accuracy при обучние: 91.111 %, 95.556 %, 97.778 % и 91.111 %;
- Class. Accuracy при тестване: 93.333 %, 93.333 %, 93.333 % и 86.667 %;
- Overall Class. Accuracy: 91.667 %, 95.000 %, 96.667 % и 90.000 %.

С най-висока ефективност е оценен Nu-SVM класификатор при “RBF” (96.667 %), докато най-ниската класификационна точност (90.000 %) е регистрирана при “Sigmoid” Kernel функция.

5. КАТЕГОРИЗАЦИЯ НА РИСК С ПОМОЩТА НА FFNNS, CFNNS И GRNNS В MATLAB

Следващите изследвания са насочени към подбор на FFNNs, CFNNs и PNNs. Реализирано е последователно обучение на FFNN и CFNN модели чрез SCG и LM алгоритми. Спецификата на тестовите трислойни невронни типове предвижда използване на:

- “tansig” в скрития и “softmax” активация в изходните слоеве;
- “tansig” в междинния и “linear” активация за изходния слой. Тяхната функционалност е оценена на база на критериите „Accuracy“ и “Cross-Entropy” за FFNNs, докато по отношение на CFNNs са въведени „Accuracy“ и “Mean-Squared Error (MSE)”. Нивата на посочените метрики за експерименталния диапазон от скрити неврони „100 до 175“ са отчетени в Таблица 2.

Относно FFNN моделите е установена най-ниска точност 51.7 % за структура с наличие на 155 междинни неврона. Максимален праг на точност 95.0 % е достигнат при задаване на 100, 125, 160 и 170 скрити неврона. По отношение на указаните невронни структури са отчени съответни нива на метрика “Cross-Entropy”: 6.91140e-0, 1.90958e-0, 3.90307e-0 и 3.09899e-0. За целите на класификационната задача е избрана архитектура с най-ниска между посочените стойности (“Cross-Entropy = 1.90958e-0”) при включване на 125 скрити неврона, дадена на Фиг. 4.а).

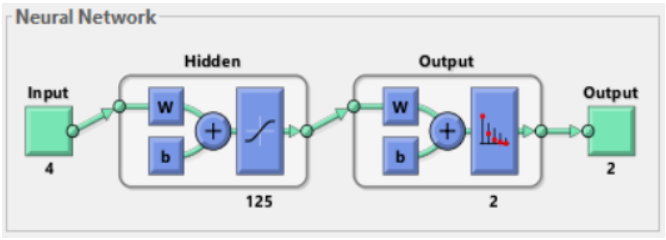
Във връзка с CFNNs е установена минимална класификационна точност 88.3 % за случаите с фискиране на 130 и 160 междинни неврона. Констатирано е, че най-високият отчетен праг на разпознаване превишава този при създадените FFNNs, като се равнява 96.7 % за невронна мрежа при задаване на 120 скрити неврона (Фиг. 4.б)). Спрямо посоченият FFNN модел е регистрирана най-ниска MSE, попадаща в порядък „0.1190“. Най-високият праг “MSE = 0.3501“ е получен за CFNN при наличие на 170 междинни неврона.

На Фиг. 5 е представена диаграма на грешките за селектираните крайни CFNN при включване на и 120 неврона в скрития слой. По отношение на синтезираната FFNN архитектура за сравнително голям дял на наблюденията за при „обучение“, „валидиране“ и „тестване“ е отчетен граничен диапазон „±0.044“. При Cascade-Forward модела са наблюдавани положителни и отрицателни вариции на мрежовите грешки в приемливи граници, въпреки наличието на изявени пикови стойности за част от наблюденията, участващи в тестовия информационен поднабор.

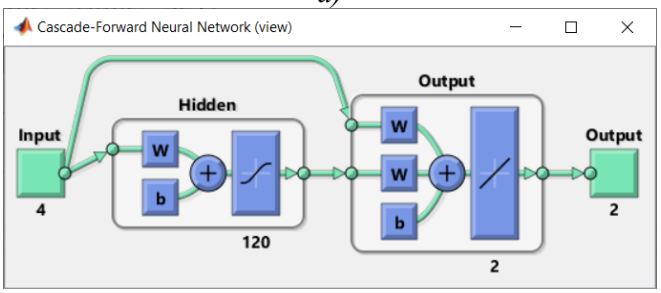
Констатираният факт за ясно изразени пикови стойности на метрика MSE може

Таблица 2. *Изследване на FFNN при SCG и CFNN модели при LM обучаващ алгоритъм*

Скрити невро-ни	*Точносту, %*	*Cross-Entropy*	*Точносту, %*	*MSE индикатор*
	FFNNs		**CFNNs**	
100	95.0	6.91140e-0	95.0	0.1258
105		7.64173e-0		0.1426
110	88.3	2.32786e-0	90.0	0.1696
115		6.83238e-0		0.1724
120	90.0	3.55443e-0	96.7	0.1190
125	95.0	1.90958e-0	93.3	0.2430
130	91.7	2.10287e-0	88.3	0.2234
135	90.0	2.27105e-0	95.5	0.2289
140	88.3	3.08918e-0	93.3	0.2741
145	93.3	3.23594e-0	91.7	0.2284
150	88.3	8.07095e-0	90.0	0.2152
155	51.7	2.27648e-0		0.1572
160	95.0	3.90307e-0	88.3	0.2079
165	88.3	7.61410e-0	91.7	0.3501
170	95.0	3.09899e-0	93.3	0.3156
175	81.7	2.67383e-0		0.2036

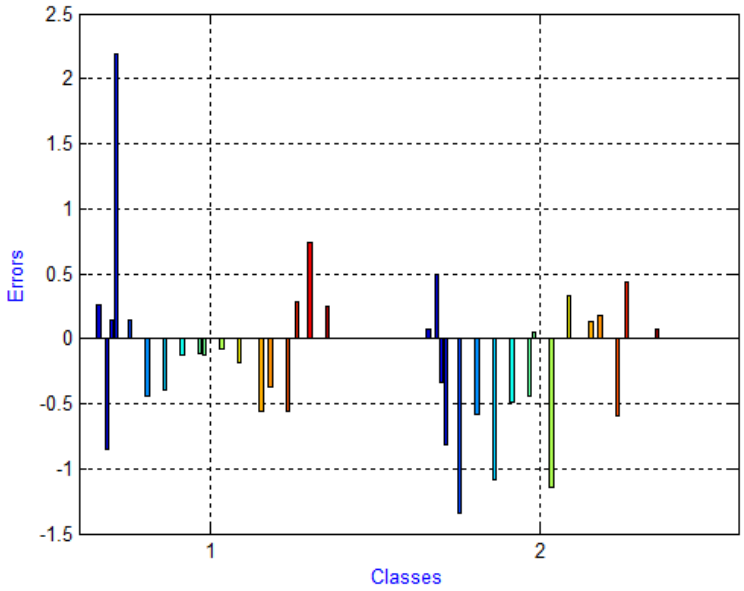


а)



б)

Фиг. 4*. Селектирани а) FFNN при SCG обучение и б) CFNN при LM алгоритъм*



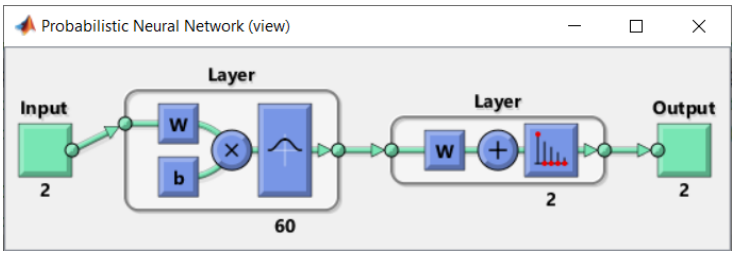
Фиг. 5*. Вариации на грешките за синтезираната CFNN архитектура при LM алгоритъм*

значително да бъде минимизиран при въвеждане на четирислойни Probabilistic Neural Networks и ограничаване на input dataset to “MVI” и “IR” входни променливи, за които е приложена обща фаза на създаване и обучение за разлика от FFNNs and CFNNs.

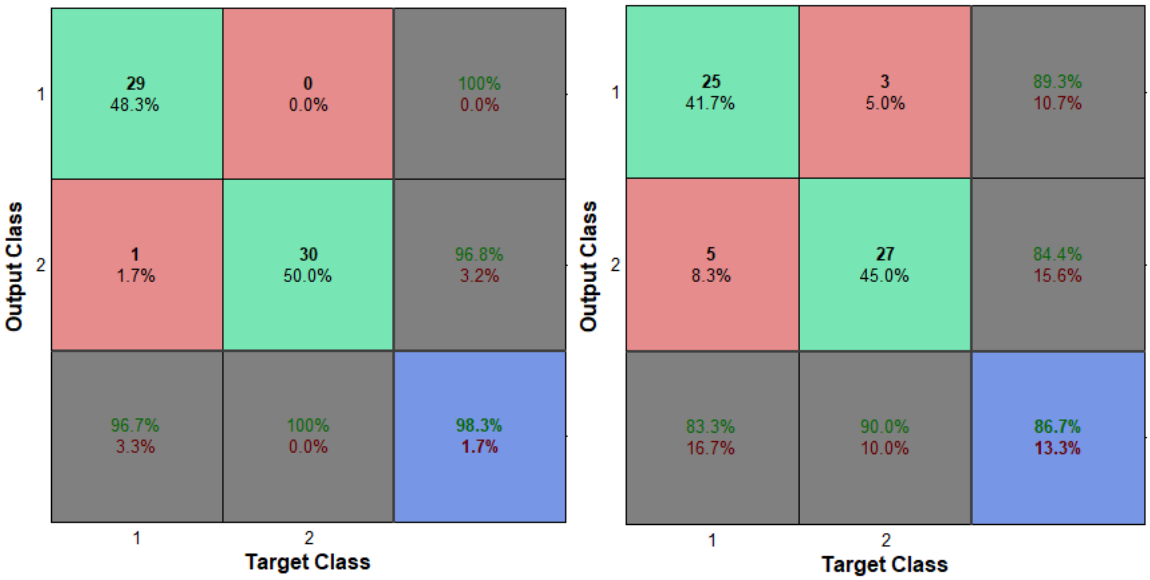
Налице е понижаване на MSE до най-висока регистрирана стойност „0.1313“ при гранична стойност на специфична метрика за функцията в радиално-базисния структурен слой “Spread = 0.890”. Най-ниските нива на индикатора са сведени до в пъти по ниската отчетена

Таблица 3. *Изследване на PNN модели при “MVI” и “IR” входни променливи*

Spread indicator	Accuracy, %	Mean-Squared Error
0.407	98.3	0.0167
0.450	95.0	0.0500
0.490		
0.530		
0.570		
0.610		
0.650	91.7	0.0833
0.690		
0.730		
0.770	90.0	0.1000
0.810	88.3	0.1167
0.850		
0.890	86.7	0.1333



а)



б) *в)*

Фиг. 6. *GRNN архитектура а) и класификационни матрици за синтезирания невронен модел при Spread нива „0.407“ б) и „0.890“ в)*

стойност “MSE = 0.0167” при начална експериментална степен на “Spread индикатор = 0.407”. Наблюдавана е тенденция на плавно намаляване на класификационната точност от максимална 98.3 % при “Spread = 0.407” до минимален праг 86.7 % за “Spread индекс = 0.890”. Изследваната концептуална PNN архитектура е показана на Фиг. 6.а), докато изходните матрици на коректни и некоректни класификации за модела с регистрирана най-висока и най-ниска точност са дадени на Фиг. 6.б) и Фиг. 6.в). С оглед на изложените резултати PNN е оценена с на-висока степен на ефективност за целите на поставената класификационна задача.

6. ЗАКЛЮЧЕНИЕ

В следващ етап от изследванията е предвидено разширяване на аналитичните апарати за разпознаване и класификация, както и интегриране на модули за апроксимация на функции. Предложеният подход се характеризира с възможности за потенциално внедряване в Business Intelligence Systems. Посочената интеграция предвижда обслужване на инвестиционни планове за подобряване на дигитализацията в информационните екосистеми на регионалните образователни центрове.

БЛАГОДАРНОСТИ: *Публикацията е във връзка с проект №BG16RFPR002-1.014-0013-C01 „Дигитализация на икономиката в среда на Големи данни– Втори етап“ (ДИГД2), финансиран по програма „Научни изследвания, иновации и дигитализация за интелигентна трансформация“ 2021 – 2027 и съфинансиран от Европейския съюз.*

ЛИТЕРАТУРА

[1] Wisniewski P, Kluza K, and Ligeza A. An approach to participatory business process modelling: BPMN model generation using constant programming and graph composition. Applied Science 2018;8(1):1-22.

[2] Ekelof L. A multiple linear regression model yo assess the effects of macroeconomic factors on small and medium-sized enterprises. Royal Institute of Technology School of Engineering Sciences 2019;1(1):1-62.

[3] Rusov J, Misita M, Milanovic D, Milanovic D. Applying regression models to predict business results. FME Transaction 2017;45(1):198-202.

[4] Dumitrescu E, Hue S, Hurlin Ch, Tokpavi S. Machine learning or econometrics for credit scoring: let’s get the best of both worlds. HAL Archives-Ouvertes Journal 2021;1(1):1-43.

[5] Yakymova L, Kuz V. The use of discriminant analysis in the assessment of municipal company's financial health. Economics and Sociology 2019;12(2):64-78.

[6] Safe M, Barouni M, Saif S. Health impact on economy by artificial neural network and dynamic ordinary least squares. Journal of Health Managements & Informatics 2017;4(4):107-113.

[7] Salleh M, Talpur N, Hussian H. Adaptive neuro-fuzzy inference system: Overview, strengths, limitations, and solutions. In: Proceedings of the international conference on data mining and big data, vol. I, 2017, p. 1-8.

[8] Hussian H, Slusarczyk B, Kamarudin F, Thaker H, Szczepanska-Woszczyna K. An investigation of an adaptive neuro-fuzzy inference system to predict the relationship among energy intensity, globalization, and financial development in major asean economies. MDPI Energies 2020;13(850):1-17.