

DOI: [10.26354/bb.4.2.103.2026](https://doi.org/10.26354/bb.4.2.103.2026)

Krzysztof Iwanicki*

ORCID: 0000-0002-3322-2756

krzysztof.iwanicki@oirpwarszawa.pl

Uczenie maszynowe w procesie walidacji modeli IRB. Regulacyjne aspekty wykorzystania z perspektywy stanowisk EBA, ECB i KNF

STRESZCZENIE

Celem artykułu jest określenie zakresu dopuszczalnego wykorzystania uczenia maszynowego w procesie walidacji modeli stosowanych w metodzie wewnętrznych ratingów (IRB). Po wyjaśnieniu istoty modeli IRB i procesu ich walidacji scharakteryzowano regulacyjne aspekty wykorzystania uczenia maszynowego w tym procesie, ze szczególnym uwzględnieniem tworzenia modeli konkurujących. Omówiono też ryzyka i ograniczenia związane z wykorzystaniem modeli uczenia maszynowego, w tym możliwość pogorszenia ich mocy predykcyjnej wskutek dryfu pojęcia oraz dryfu danych. Kolejno przedmiotem rozważań jest znaczenie wyjaśnialności w procesie walidacji. W konkluzji stwierdza się, że z obowiązujących regulacji wynika dopuszczalność wykorzystania uczenia maszynowego jako jednego z narzędzi stosowanych w procesie walidacji modeli IRB. Jednocześnie obecny stan wiedzy z zakresu informatyki nie daje podstaw do przyjęcia, że uczenie maszynowe może samodzielnie zapewnić realizację wszystkich celów walidacji. W konsekwencji wykorzystanie uczenia maszynowego nie może stanowić funkcjonalnego substytutu procesu walidacji.

Słowa kluczowe: modele IRB, walidacja modeli IRB, uczenie maszynowe, ryzyko modelu, wyjaśnialna sztuczna inteligencja (XAI)

Kody JEL: G21, G28, C52, C45, C53

* Krzysztof Iwanicki – Okręgowa Izba Radców Prawnych w Warszawie.

MACHINE LEARNING IN THE VALIDATION OF IRB MODELS: REGULATORY ASPECTS OF ITS USE IN THE LIGHT OF THE POSITIONS OF THE EBA, ECB AND KNF

ABSTRACT

The aim of the article is to define the scope of acceptable use of machine learning in the process of validating models used in the internal ratings-based (IRB) approach. After clarifying the essence of IRB models and their validation process, the regulatory aspects of using machine learning in this process are described, with particular attention to the creation of competing models. The risks and limitations associated with using machine learning models are also discussed, including the possibility of reduced predictive power due to concept drift and data drift. The discussion also focuses on the importance of explainability in the validation process. In conclusion, it is stated that current regulations allow the use of machine learning as one of the tools in the IRB model validation process. At the same time, the current state of knowledge in computer science does not provide a basis for assuming that machine learning alone can fully achieve all validation goals. As a result, using machine learning can't really replace the validation process.

Keywords: IRB models, IRB model validation, machine learning, model risk, explainable artificial intelligence (XAI)

JEL Codes: G21, G28, C52, C45, C53

1. MODELE IRB

Ryzyko nieodłącznie towarzyszy działalności bankowej. Wynika ono z pełnionej przez banki roli pośredników finansowych (Zawadzka 2001, s. 52; Matuszyk 2008, s. 13); działalność bankowa polega m.in. na podejmowaniu ryzyka i zarządzania nim (Iwanicz-Drozdowska 2024, s. 17). Ryzyko definiuje się jako: „*możliwość nieosiągnięcia zamierzonego celu (efektu), czyli odchylenie in minus od stanu pożądanego (negatywne podejście do ryzyka) lub też każde odchylenie (zarówno in plus, jak i in minus) od zamierzonego celu (podejście neutralne)*” (Iwanicz-Drozdowska, Chmielewski 2024, s. 43), albo jako zobiiektywizowaną niepewność, opartą na mierzalnych przesłankach, co odróżnia je od niepewności (Niedziółka 2018, s. 305–306). Na potrzeby niniejszego opracowania przyjęto pierwszą definicję.

W ramach ryzyka bankowego wyodrębnia się ryzyko kredytowe. W literaturze przedmiotu wyróżnia się wąskie i szerokie ujęcie tej kategorii. W wąskim ujęciu znaczy ono: „*niebezpieczeństwo niespłacenia przez dłużnika banku [...] zaciągniętego kredytu wraz z odsetkami i prowizjami*” (Matuszyk, Ptak-Chmielewska 2024, s. 67), zaś w ujęciu szerokim: „*ryzyko niewywiązywania się ze swoich zobowiązań partnera transakcji wobec banku i dotyczy kredytów, gwarancji, poręczeń, akceptów bankowych, akredytyw, lokat międzybankowych, instrumentów dłużnych –*

pozagiełdowych instrumentów pochodnych” (Matuszyk, Ptak-Chmielewska 2024, s. 67–68). Dalsze analizy oparto na pierwszym ujęciu.

Jednym z podstawowych elementów zarządzania indywidualnym ryzykiem kredytowym jest dokonywanie oceny zdolności kredytowej, czyli: *„zdolności do terminowego i kompletnego wypełniania zobowiązań oraz warunków umowy kredytowej”* (Wiatr 2008, s. 77). Ocena ta jest dokonywana m.in. poprzez wykorzystanie modeli, które upraszczają rzeczywistość: *„poprzez użycie metod ilościowych oraz eksperckich”* (Komisja Nadzoru Finansowego 2024, s. 13).

Pomiar ryzyka kredytowego, w tym obliczanie kwot ekspozycji ważonych ryzykiem, może odbywać się z wykorzystaniem bądź podejścia standardowego, którego reguły są określane przez przepisy i nadzorcę, bądź metody IRB (ang. *Internal Ratings-Based*), to jest metody wewnętrznych ratingów, która opiera się na założeniu, że: *„bank jest w stanie sam odpowiednio precyzyjnie ocenić rozmiary podejmowanego ryzyka kredytowego. Istotą podejścia IRB jest możliwość wykorzystania wewnętrznie oszacowanych parametrów ryzyka”* (Schab 2005, s. 93). Podstawą prawną posługiwania się metodą IRB jest art. 143 ust. 1 rozporządzenia Parlamentu Europejskiego i Rady (UE) nr 575/2013 z dnia 26 czerwca 2013 r. w sprawie wymogów ostrożnościowych dla instytucji kredytowych i firm inwestycyjnych, zmieniające rozporządzenie (UE) nr 648/2012 (Dz. U. UE. L. z 2013 r. Nr 176, str. 1 z późn. zm.), dalej „CRR”, zgodnie z którym właściwy organ zezwala instytucjom na obliczanie kwot ekspozycji ważonych ryzykiem przy wykorzystaniu metody IRB po spełnieniu określonych warunków.

2. WALIDACJA MODELI IRB

Stosowanie IRB jest związane z obowiązkiem walidacji, czyli dokonywania *„oceny skuteczności działania modelu”* (Komisja Nadzoru Finansowego 2015, s. 8).

ECB wskazuje, że zadaniem walidacji jest: *„sformułować opinię, czy finalny system ratingowy [...] spełnia wymagania regulacyjne”* (European Central Bank 2023, s. 15)¹ oraz: *„zapewnić skuteczny challenge adekwatności działania modelu i odpowiedniości systemu ratingowego”* (European Central Bank 2023 s. 16)². Z kolei KNF wskazuje, że: *„podstawowym celem walidacji modeli [...] jest redukcja poziomu ryzyka modeli poprzez: weryfikację poprawności i ocenę skuteczności działania modelu oraz sposobu jego funkcjonowania, identyfikację obszarów, w których można wprowadzić uzasadnione udoskonalenia dotyczące zarówno w sposób bezpośredni samego modelu, jak i procesu jego stosowania, formułowanie odpowiednich zaleceń i nadawanie im odpowiedniej priorytetyzacji, weryfikację jakości realizacji sformułowanych zaleceń, promowanie wysokich standardów zarządzania modelami”* (Komisja Nadzoru Finansowego 2015, s. 39).

¹ Przekład własny, w oryginale: *„form an opinion on whether the final rating system [...] meets the regulatory requirements”* (European Central Bank 2023, s. 15).

² Przekład własny, w oryginale: *„ensuring an effective challenge for the adequate model performance and appropriateness of the rating system”* (European Central Bank 2023, s. 16).

ECB podaje, że w procesie walidacji wykorzystuje się między innymi dwa rodzaje prób. Pierwszą jest „próba metryk efektywności” (ang. *performance metric samples*), która polega na takim przetestowaniu modelu, aby: „*dłużnicy lub transakcje były przypisywane do klas lub puli, a metryki empiryczne (przykładowo niewykonanie zobowiązania) były obserwowane, w zakresie, w jakim jest to możliwe, do oceny podstawowej efektywności modelu*” (European Central Bank 2023, s. 66)³. Drugą jest „próba testowego zastosowania” (ang. *application test samples*), która polega na tym, aby: „*dłużników lub transakcje przypisać do klas lub puli w sytuacji, w której nie upłynęło wystarczająco dużo czasu, aby zaobserwować realizacje empiryczne potrzebne do oceny modelu*” (European Central Bank 2023, s. 66)⁴. W praktyce próba ta polega na przetestowaniu modelu opartego na aktualnych ekspozycjach, które nie są jeszcze wymagalne.

3. WPROWADZENIE DO WYKORZYSTANIA UCZENIA MASZYNOWEGO

Możliwość wykorzystania uczenia maszynowego w procesie walidacji modeli IRB była wskazywana w dokumentach EBA co najmniej od 2021 r. W tym okresie EBA rozpoczęła konsultacje wśród europejskich podmiotów sektora finansowego, wydając dokument, w którym stwierdziła, że: „*głównymi zidentyfikowanymi płaszczyznami wykorzystania [uczenia maszynowego są – przyp. KI] [...] walidacja modeli, gdzie uczenie maszynowe może być użyte, na przykład, jako model konkurujący lub do benchmarkingu*” (European Banking Authority 2021, s. 13)⁵. W analizowanym okresie wykorzystanie uczenia maszynowego w szerokim odniesieniu do modeli IRB było traktowane z ostrożnością (European Banking Authority 2021, s. 10), a: „*instytucje podjęły decyzję o odejściu od użycia uczenia maszynowego w obszarach regulowanych*” (European Banking Authority 2021, s. 11)⁶.

Istnieje różnica między oparciem modelu IRB na uczeniu maszynowym a wykorzystaniem uczenia maszynowego do walidacji modelu IRB nieopartego na uczeniu maszynowym. Różnica ta jest istotna, choć nie zawsze wyartykułowana *expressis verbis*. Dlatego wyprowadzenie tezy z danego źródła wymaga wprawdzie zbadania, w jakim zakresie źródło odnosi się do pierwszego ujęcia, a w jakim do drugiego. Przykładowo, publikacja ECB pod tytułem „*Przewodnik po modelach wewnętrznych*” analizom w zakresie wykorzystania uczenia maszynowego w odniesieniu do modeli wewnętrznych poświęca cały rozdział, jednakże analizy te odnoszą się do modeli wewnętrznych opartych na uczeniu maszynowym (European Central Bank 2025, s. 15–26), toteż użyteczność tego źródła dla niniejszego opracowania jest ograniczona.

³ Przekład własny, w oryginale: *obligors or facilities are assigned to grades or pools and the empirical metrics (e.g. defaults) can be observed to assess the core model performance* (European Central Bank 2023, s. 66).

⁴ Przekład własny, w oryginale: „*obligors or facilities are assigned to grades or pools, but there has not been enough time to observe the empirical realisations to assess the model*” (European Central Bank 2023, s. 66).

⁵ Przekład własny, w oryginale: „*The identified main areas of use are: [...] Model validation, where ML models can be used, for example, as model challengers or for benchmarking*” (European Banking Authority 2021, s. 13).

⁶ Przekład własny, w oryginale: „*institutions have made the strategic decision to move away from using ML in regulatory areas*” (European Banking Authority 2021, s. 11).

EBA w dokumencie rozpoczynającym konsultacje skonkretyzowała zasadnicze obszary wykorzystania uczenia maszynowego, wskazując na: (a) różnicowanie ryzyka, (b) kwantyfikację ryzyka, (c) walidację modeli. Pozostałe obszary dodatkowo skategoryzowała zbiorczym punktem „pozostałych obszarów”. Pamiętając o uwagach nakreślonych w akapicie poprzedzającym, należy stwierdzić, że obszary (a) i (b) odnoszą się do wykorzystania uczenia maszynowego w modelu (model IRB oparty na uczeniu maszynowym), a z kolei obszar (c) odnosi się do jego wykorzystania w procesie walidacji (model IRB niekoniecznie musi być oparty na uczeniu maszynowym).

W raporcie podsumowującym konsultacje nie sformułowano obszarów daleko odbiegających od wskazanych na wstępie; w odniesieniu do walidacji doprecyzowano, że: „*uczenie maszynowe może być wykorzystywane [...] do celów walidacji wewnętrznej*” (European Banking Authority 2023, s. 20)⁷, oraz, że uczenie maszynowe można wykorzystać, aby: „*budować modele challenger [...] oraz prowadzić benchmarking w walidacji początkowej i bieżącej*” (European Banking Authority 2023, s. 7)⁸, czy też w zakresie badania generalizacji: „*z wykorzystaniem prób danych out-of-sample i out-of-time*” (European Banking Authority 2023, s. 8)⁹.

4. DOPUSZCZALNOŚĆ WYKORZYSTANIA UCZENIA MASZYNOWEGO

Przepisy nie ustanawiają zamkniętego katalogu technik, które mogą być wykorzystane w procesie walidacji. Art. 185 lit. c CRR posługuje się szerokim pojęciem „innych narzędzi walidacji”, które ma charakter otwarty. Analogiczne podejście przyjęto w rozporządzeniu delegowanym Komisji (UE) 2022/439 z dnia 20 października 2021 r. *uzupełniającym rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 575/2013 w odniesieniu do regulacyjnych standardów technicznych dotyczących określenia metodyki oceny, którą właściwe organy mają stosować podczas oceny spełniania przez instytucje kredytowe i firmy inwestycyjne wymogów stosowania metody wewnętrznych ratingów* (Dz. U. UE. L. z 2022 r. Nr 90, str. 1), dalej „rozporządzenie 2022/439”, w którym przepis art. 12 lit. b, odnoszący się między innymi do adekwatności metod walidacji, wskazuje, że: „*metody [...] walidacji są odpowiednie do charakteru, stopnia złożoności i zakresu stosowania systemów ratingowych instytucji oraz dostępności danych*”. Ten przepis również ma charakter otwarty.

⁷ Przekład własny, w oryginale: „*ML might be used [...] for internal validation purposes*” (European Banking Authority 2023, s. 20).

⁸ Przekład własny, w oryginale: „*develop challenger models [...] and for benchmarking purposes in the initial as well as ongoing validation*” (European Banking Authority 2023, s. 7).

⁹ Przekład własny, w oryginale: „*using out-of-sample and out-of-time data sample*” (European Banking Authority 2023, s. 8).

Taka wykładnia jest zgodna ze stanowiskiem ECB, który: „oczekuje, że zostaną zastosowane wszystkie odpowiednie techniki walidacji” (European Central Bank 2023, s. 23–24)¹⁰. Co więcej, stanowisko ECB idzie przy tym dalej i oczekuje, że: „instytucje będą poszukiwać najnowocześniejszych technik walidacji” (European Central Bank 2023, s. 25)¹¹.

5. WYKORZYSTANIE UCZENIA MASZYNOWEGO PRZY TWORZENIU MODELI KONKURUJĄCYCH

Modele IRB, które są walidowane między innymi z wykorzystaniem próby metryk efektywności, opierają się na danych historycznych. Dane te, nawet jeżeli ich prawdziwość jest niekwestionowana, mogą nie odzwierciedlać aktualnych warunków, które mogły się zmienić wskutek zmiany uwarunkowań rynkowych (przemian ekonomicznych, nowelizacji otoczenia regulacyjnego, redefiniowania polityki kredytowej kredytodawcy, zmian w zachowaniu kredytobiorców, etc.). Narzędziami pozwalającymi identyfikować ten stan rzeczy są modele konkurujące (ang. *challenger models*). Porównanie wyników walidowanego modelu IRB z wynikami modelu konkurującego pozwala ocenić, czy model IRB podlegający walidacji adekwatnie odwzorowuje zależności występujące w bieżącym portfelu. Jeżeli model konkurujący osiąga lepsze wyniki w zakresie trafności predykcji, może to świadczyć o tym, że walidowany model IRB nie przystaje do obecnych uwarunkowań rynkowych. W tych okolicznościach model konkurujący nie zastępuje modelu IRB, lecz pełni funkcję benchmarku.

Jak wskazuje KNF: „metody uczenia maszynowego odznaczają się wyższą (najczęściej zauważalnie wyższą) siłą dyskryminacyjną od standardowej metody regresji logistycznej” (Komisja Nadzoru Finansowego 2024, s. 70), co przemawia za ich użytecznością, tym bardziej, że: „uczenie maszynowe ma charakter bardziej empiryczny [...] w tym uwzględnia interakcje wyższego rzędu, które nie są z góry określone” (Komisja Nadzoru Finansowego 2026, s. 62). Jak wynika z konsultacji przeprowadzonych przez ECB, uczenie maszynowe przede wszystkim było wykorzystywane do opracowywania modeli obliczających PD; do obliczania LGD, EAD, ELBE i CCF wykorzystywano je rzadziej (European Central Bank 2023, s. 6).

6. WYKORZYSTANIE UCZENIA MASZYNOWEGO W OBSZARZE ZARZĄDZANIA DANymi

Drugim obszarem wykorzystania uczenia maszynowego jest wsparcie analizy jakości danych. W poprzednim rozdziale założono, że dane statystyczne są prawdziwe, w niniejszym zakłada się, że zawierają błędy charakterystyczne dla baz statystycznych.

¹⁰ Przekład własny, w oryginale: „all relevant validation techniques are expected to be used” (European Central Bank 2023 s. 23–24).

¹¹ Przekład własny, w oryginale: „institutions are expected to seek for state-of-the art validation techniques” (European Central Bank 2023, s. 25).

Jak wynika z rekomendacji W: „[w] odniesieniu do każdego ze stosowanych modeli istotnych bank powinien oceniać ewentualne narażenie modelu na [...] ryzyko danych wynikające z [...] niskiej jakości danych wejściowych; wadliwości procesów pozyskiwania, przetwarzania, agregowania i składowania danych; niewystarczającej liczebności próby; nieadekwatnej długości szeregów czasowych; nieodpowiedniej częstotliwości zasilania modelu danymi; niewystarczalności dostępnego zakresu danych w kontekście opisu przebiegu” (Komisja Nadzoru Finansowego 2015, s. 24).

Jak wskazuje KNF: „zastosowanie technik uczenia maszynowego [...] umożliwia efektywne wykorzystanie nie zawsze ustrukturyzowanych danych pochodzących z różnych, czasami niekonwencjonalnych, źródeł” (Komisja Nadzoru Finansowego 2026, s. 63). Jednocześnie KNF wskazuje w tym samym dokumencie na związane z tym ryzyko, bowiem: „niezbilansowane zestawy danych [...] W takich sytuacjach modele uczenia maszynowego mogą działać niepoprawnie” (Komisja Nadzoru Finansowego 2026, s. 70–71).

7. RYZYKO NIEWYJAŚNIALNOŚCI

Zgodnie z art. 175 ust. 4 CRR: „jeżeli instytucja korzysta z modeli statystycznych w ramach procesu ratingu, dokumentuje ona stosowane w tym względzie metody. Dokumentacja ta: (a) zawiera szczegółowy opis teorii, założeń oraz podstaw matematycznych i empirycznych”. W niektórych przypadkach wymagane jest dodatkowe spełnienie wymogu wynikającego z art. 3 ust. 1 lit. d rozporządzenia 2022/439, który stanowi, że dokumentacja systemów ratingowych musi zezwalać: „osobom trzecim na zbadanie i potwierdzenie funkcjonowania systemów ratingowych oraz, w szczególności, na zbadanie i potwierdzenie, że: [...] (i) dokumentacja konstrukcji systemu ratingowego jest wystarczająco szczegółowa, by umożliwić osobie trzeciej zrozumienie rozumowania stojącego za wszystkimi aspektami systemu ratingowego”. W tym kontekście rodzi się pytanie, czy walidacja dokonana z użyciem uczenia maszynowego będzie spełniać powyższy wymóg (w dalszej części opracowania nazwany będzie skrótowo jako „wymóg wystarczającej szczegółowości”).

Nie sposób ustalić, czy uczenie maszynowe: „umożliwia zrozumienie rozumowania stojącego za wszystkimi aspektami systemu ratingowego” bez odwołania się do nauk informatycznych. Jak wskazuje Ch. Molnar, modele wyjaśnialne mogą być zrozumiane jedynie z wykorzystaniem dodatkowych technik, podczas gdy do zrozumienia modeli interpretowalnych wystarczające jest jedynie poznanie ich budowy. Przyjmując tę perspektywę pojęcie interpretowalności, należy rozumieć jako stopień, w jakim człowiek może zrozumieć powody, dla których dany system wygenerował dany wynik; alternatywnie interpretowalność można jako stopień, w jakim człowiek może przewidzieć wynik (Molnar 2019, s. 15). Systemy uczenia maszynowego można sklasyfikować według stopnia ich interpretowalności, poczynawszy od tych z niską interpretowalnością, skończywszy na tych wysoce interpretowalnych. Jak trafnie zauważa M. Krzysztoń: „w przypadku systemów trudniejszych w interpretacji, gdy wymagane jest zrozumienie sposobu

ich działania, konieczne jest wykorzystanie dodatkowych narzędzi z obszaru XAI (ang. explainable AI). Standardowym podejściem jest wyjaśnialność post hoc, tj. po wytrenowaniu modelu” (Krzysztoń 2023, s. 52).

Warto rozważyć, czy wskazany powyżej wymóg prawny, dotyczący możliwości: „zrozumienia rozumowania stojącego za wszystkimi aspektami systemu ratingowego” należy utożsamiać z interpretowalnością, czy z wyjaśnialnością. Rozważyć także należy, czy i pod jakimi warunkami wyjaśnienie uzyskane z pomocą narzędzi XAI spełni wymóg wystarczającej szczegółowości. W analizowanych dokumentach EBA, ECB i KNF kwestia ta nie została rozstrzygnięta. Tytułem przykładu można przywołać stanowisko KNF, zgodnie z którym: „model jest możliwy do wyjaśnienia, jeśli jego wewnętrzne działanie może być zrozumiane przez ludzi albo kiedy możliwe jest uzasadnienie wyniku modelu” (Komisja Nadzoru Finansowego 2026, s. 72), to jednak *explicite* nie precyzuje, czy „uzasadnienie wyniku modelu”, w rozumieniu stanowiska zaprezentowanego w publikacji wydanej nakładem UKNF, może polegać na przedstawieniu wyjaśnienia wygenerowanego przez narzędzie XAI ani jakie wymagania takie wyjaśnienie musiałoby spełniać¹².

W tym jednak miejscu w ślad za EBA warto zwrócić uwagę, że przy tworzeniu systemów opartych na uczeniu maszynowym należy: „unikać niepotrzebnej złożoności [...], jeśli nie uzasadnia jej istotna poprawa zdolności predykcyjnych” (European Banking Authority 2023, s. 21), aby nie utrudniać badania wyjaśnialności tych systemów.

8. RYZYKO POGORSZENIA ZDOLNOŚCI PREDYKCYJNEJ

W literaturze szeroko traktuje się o ryzykach wykorzystania uczenia maszynowego, skutkujących zdegradowaniem systemu (Szostek i in. 2023, s. 25). Jednak część podatności na degradację zdaje się nie występować w systemach wykorzystywanych w procesie walidacji modeli IRB. Wynika to między innymi z tego, że systemy te nie jawią się jako atrakcyjne dla potencjalnych atakujących. Mają one bowiem charakter pośredni i nie oddziałują bezpośrednio na decyzję kredytową, dlatego nie dają w zasadzie atakującemu sposobności monetyzacji swojej działalności. Jednakże ryzyka związane z atakami stanowią odrębny problem, a z uwagi na zakres niniejszego opracowania dalsze analizy ograniczono do ryzyk wynikających z właściwości samych modeli.

Pogorszenie zdolności predykcyjnej może nastąpić bez ingerencji osób trzecich. Trafne analizy w tym zakresie przeprowadził J. Surma, który, odwołując się do definicji T. Mitchella, wskazał, że: „system jest się w stanie uczyć z doświadczenia (ang. experience) *E* w kontekście realizacji zadań (ang. tasks) *T* i miary jakości działania (ang. performance measure) *P*, jeśli jego działanie w realizacji zadań *T*, mierzone z wykorzystaniem *P*, polepsza się wraz ze wzrostem doświadczenia *E*. W efekcie procesu uczenia się system jest w stanie generować poprawną odpowiedź

¹² Z uwagi na ograniczenia objętości artykułu nie sposób przesądzić tych kwestii w ramach niniejszego opracowania. Zasygnalizowane w tym fragmencie problemy będą przedmiotem dalszych badań, których wyniki zostaną przedstawione w odrębnych opracowaniach.

(wyjście) dla danego obiektu na wejściu” (Surma 2023, s. 36). Sformułował on potencjalne skutki wadliwego działania takiego systemu, w tym obniżenie jakości klasyfikatora poprzez: „generowanie błędów fałszywie pozytywnych lub fałszywie negatywnych. Konsekwencją tego typu ataku jest spadek dokładności klasyfikacji, co implikuje obniżenie wiarygodności systemu (ang. confidence reduction), a nawet, w sytuacji skrajnej, rezygnację z jego użytkowania. Wynika to m.in. z faktu, że błędne klasyfikacje generują realne i potencjalne (np. związane z utratą reputacji) koszty będące konsekwencją błędnych decyzji albo ich braku” (Surma 2023, s. 36–37).

Degradacja jakości klasyfikatora może nastąpić bez ingerencji człowieka. Wśród opisywanych w literaturze przyczyn zaistnienia tego zjawiska podkreśla się dwa następujące. Po pierwsze, może nastąpić „dryf danych” (ang. *data drift*), oznaczający zmianę rozkładu danych wejściowych, która nie musi wiązać się ze zmianą zależności między danymi wejściowymi a zmienną objaśnianą (Gama i in. 2014, s. 44:4). Innymi słowy, model, który nauczył się działać na podstawie zbioru danych historycznych (choćaby takich, jakie są wykorzystywane do próby efektywności metryk) nie działa równie dobrze na innym zbiorze danych. Po drugie, może zaistnieć „dryf pojęcia” (ang. *concept drift*), polegający na zmianie zależności między zmiennymi wejściowymi a zmienną objaśnianą (Gama i in. 2014, s. 44:4). W przypadku walidacji modeli IRB może oznaczać to, że oceniany parametr ma w aktualnych warunkach inne znaczenie predykcyjne aniżeli w przeszłości (przykładowo, gdy wysoka kwota kredytu historycznie mogła stanowić o wzroście prawdopodobieństwa niewykonania zobowiązania, obecnie może wynikać z podwyższonej inflacji i zmian stóp procentowych).

PODSUMOWANIE

Przeprowadzone analizy na styku dyscypliny nauk prawnych z dyscypliną nauk informatycznych doprowadziły do następujących konkluzji w przedmiocie zakresu wykorzystania uczenia maszynowego w procesie walidacji modeli IRB.

1. Z analizy przepisów CRR i rozporządzenia 2022/439, dokonanej z uwzględnieniem stanowisk EBA, ECB i KNF, wynika dopuszczalność wykorzystania uczenia maszynowego jako jednego z narzędzi stosowanych w procesie walidacji modeli IRB.
2. Obecny stan wiedzy z zakresu informatyki nie daje podstaw do przyjęcia, że wykorzystanie uczenia maszynowego może samodzielnie zapewnić realizację wszystkich celów walidacji modeli IRB. Modele uczenia maszynowego podlegają bowiem ograniczeniom oddziaływającym na wiarygodność ich wyników.
3. Zestawienie perspektywy normatywnej oraz informatycznej prowadzi do wniosku, że wykorzystanie uczenia maszynowego nie może stanowić funkcjonalnego substytutu walidacji.

Ponadto w toku przeprowadzonych analiz ujawniono nowe problemy, które pozwalają sformułować przynajmniej dwa pytania badawcze:

1. Czy, a jeżeli tak to pod jakimi warunkami wyjaśnienie uzyskane z pomocą narzędzi XAI spełni wymóg wystarczającej szczegółowości, o którym mowa w art. 3 ust. 1 lit. d rozporządzenia 2022/439?
2. Czy „uzasadnienie wyniku modelu”, w rozumieniu stanowiska UKNF, może polegać na przedstawieniu wyjaśnienia wygenerowanego przez narzędzie XAI?

Na zakończenie warto zaakcentować, że jak trafnie stanowi rekomendacja W: „*kształtując proces walidacji należy pamiętać, że nie ma jednego uniwersalnego podejścia do walidacji, które byłoby zawsze dobrze wkomponowane w specyfikę banku i gamę wykorzystywanych w nim modeli – dlatego sposób realizacji zadań wynikających z metodyki walidacji (a niekiedy wykraczających poza ściśle wyspecyfikowane ramy) powinien być każdorazowo elastycznie dostosowywany do charakterystyki modelu*” (Komisja Nadzoru Finansowego 2015, s. 40). Dlatego również zakres wykorzystania uczenia maszynowego powinien być każdorazowo określany z uwzględnieniem celów walidacji, charakterystyki modelu oraz specyfiki banku.

BIBLIOGRAFIA

European Banking Authority (2021), *Discussion Paper on Machine Learning for IRB Models*, https://www.eba.europa.eu/sites/default/files/document_library/Publications/Discussions/2022/Discussion%20on%20machine%20learning%20for%20IRB%20models/1023883/Discussion%20paper%20on%20machine%20learning%20for%20IRB%20models.pdf

European Banking Authority (2023), *Machine Learning for IRB Models. Follow-up Report from the Consultation on the Discussion Paper on Machine Learning for IRB Models*, https://www.eba.europa.eu/sites/default/files/document_library/Publications/Reports/2023/1061483/Follow-up%20report%20on%20machine%20learning%20for%20IRB%20models.pdf

European Central Bank (2023), *Supervisory Handbook on the Validation of IRB Rating Systems Under the Internal Ratings Based Approach*, https://www.eba.europa.eu/sites/default/files/document_library/Publications/Reports/2023/1061495/Supervisory%20handbook%20on%20the%20validation%20of%20IRB%20rating%20systems%20revised.pdf

European Central Bank (2025), *ECB guide to internal models*, https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guide202507.en.pdf

Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. (2014), *A Survey on Concept Drift Adaptation*, „ACM Computing Surveys”, t. 46, nr 4, <https://doi.org/10.1145/2523813>.

Iwanicz-Drozdowska M. (2024), *Wprowadzenie do zagadnień ryzyka bankowego*, [w:] *Zarządzanie ryzykiem bankowym*, M. Iwanicz-Drozdowska (red.), Wolters Kluwer, Warszawa.

Iwanicz-Drozdowska M., Chmielewski T. (2024), *Kluczowe hasła*, [w:] *Zarządzanie ryzykiem bankowym*, M. Iwanicz-Drozdowska (red.), Wolters Kluwer, Warszawa.

Komisja Nadzoru Finansowego (2015), *Rekomendacja W dotycząca zarządzania ryzykiem modeli w bankach*, https://www.knf.gov.pl/knf/pl/komponenty/img/knf_137749_Rekomendacja_W_42219.pdf

Komisja Nadzoru Finansowego (2024), *Zarządzanie ryzykiem modeli w zakresie działalności podmiotów sektora bankowego ze szczególnym uwzględnieniem modeli ryzyka kredytowego – podstawowe zagadnienia*, https://www.knf.gov.pl/knf/pl/komponenty/img/publikacja_zarzadzanie_ryzykiem_modeli_92496.pdf

Komisja Nadzoru Finansowego (2026), *Rekomendacje dla podmiotów rynku finansowego dotyczące wpływu modeli Frontier AI na cyberbezpieczeństwo*, https://www.knf.gov.pl/knf/pl/komponenty/img/Rekomendacje_dla_podmiotow_ryнку_finansowego_dot_wplywu_modeli_Frontier_AI_na_cyberbezpieczenstwo_99167.pdf

Krzysztoń M. (2023), *Weryfikacja wiarygodności systemów w erze uczenia maszynowego*, [w:] NASK *Cyberbezpieczeństwo w AI. AI w Cyberbezpieczeństwie*, <https://cyberpolicy.nask.pl/wp-content/uploads/2023/09/Cyberbezpieczenstwo-AI.-AI-w-cyberbezpieczenstwie.pdf>

Matuszyk A. (2008), *Credit scoring*, CeDeWu, Warszawa.

Matuszyk A., Ptak-Chmielewska A. (2024), *Ryzyko kredytowe*, [w:] *Zarządzanie ryzykiem bankowym*, M. Iwanicz-Drozdowska (red.), Wolters Kluwer, Warszawa.

Molnar Ch. (2019), *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*, https://originalstatic.aminer.cn/misc/pdf/Molnar-interpretable-machine-learning_compressed.pdf

Niedziółka P. (2018), *Pojęcie i rodzaje ryzyka bankowego*, [w:] *Świat bankowości*, M. Zaleska (red.), Difin, Warszawa.

Rozporządzenie delegowane Komisji (UE) 2022/439 z dnia 20 października 2021 r. uzupełniające rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 575/2013 w odniesieniu do regulacyjnych standardów technicznych dotyczących określenia metodyki oceny, którą właściwe organy mają stosować podczas oceny spełniania przez instytucje kredytowe i firmy inwestycyjne wymogów stosowania metody wewnętrznych ratingów, Dz.U. UE L 90 z 18.03.2022, s. 1–66.

Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 575/2013 z dnia 26 czerwca 2013 r. w sprawie wymogów ostrożnościowych dla instytucji kredytowych i firm inwestycyjnych, zmieniające rozporządzenie (UE) nr 648/2012, Dz.U. UE L 176 z 27.06.2013, s. 1–337, ze zm.

Schab I. (2005), *Ocena ryzyka kredytowego w ramach wewnętrznych systemów ratingowych – charakterystyka podejścia oraz podstawowych wymogów*, „Bezpieczny Bank”, 26(1), <https://ojs.bfg.pl/index.php/bb/article/view/641>

Surma J. (2023), *Wprowadzenie do ataków na systemy uczenia maszynowego*, [w:] *Cyberbezpieczeństwo AI. AI w cyberbezpieczeństwie*, CyberPolicy NASK, <https://cyberpolicy.nask.pl/wp-content/uploads/2023/09/Cyberbezpieczenstwo-AI.-AI-w-cyberbezpieczenstwie.pdf>

Szostek D., Kasprowski P., Kozak J., Kapczyński A., Prabucki R. (2023), *Wyzwania i zagrożenia z zakresu cyberbezpieczeństwa podczas projektowania lub wykorzystywania AI*, [w:] *Cyberbezpieczeństwo AI. AI w cyberbezpieczeństwie*, CyberPolicy NASK, <https://cyberpolicy.nask.pl/wp-content/uploads/2023/09/Cyberbezpieczenstwo-AI.-AI-w-cyberbezpieczenstwie.pdf>

Wiatr M. (2008), *Zarządzanie indywidualnym ryzykiem kredytowym. Elementy systemu*, Oficyna Wydawnicza SGH, Warszawa.

Zawadzka Z. (2001), *Ryzyko bankowe*, „Bank”, nr 6.