Krzysztof Łuczka[1]
Uniwersytet Ekonomiczny w Poznaniu
https://doi.org/10.29316/9788368103304_6

ZMIANY I INNOWACJE W OCENIE ZDOLNOŚCI KREDYTOWEJ

CHANGES AND INNOVATIONS IN CREDITWORTHINESS ASSESSMENT

Streszczenie

Celem rozdziału jest przedstawienie ewolucji metod oceny zdolności kredytowej oraz roli nowoczesnych technologii w tym procesie. Analiza obejmuje zarówno klasyczne podejścia statystyczne, jak i współczesne metody oparte na uczeniu maszynowym, sztucznej inteligencji oraz dużych modelach językowych (LLM). W pierwszej części omówiono historyczne ujęcie problemu, począwszy od wczesnych form oceny opartej na reputacji i subiektywnej analizie kredytodawców, aż po rozwój systemu punktowego FICO i standaryzację credit scoringu. Następnie zaprezentowano najnowsze rozwiązania, takie jak modele hybrydowe (np. TCN-DilateFormer), algorytmy zespołowe (Random Forest, XGBoost, TabNet) oraz zastosowanie metod głębokiego uczenia w prognozowaniu ryzyka kredytowego. Praca omawia również polskie uwarunkowania systemowe - w szczególności funkcjonowanie Biura Informacji Kredytowej oraz znaczenie Rekomendacji T w zarządzaniu ryzykiem detalicznym. Wnioski wskazują, że choć metody tradycyjne pozostają fundamentem oceny kredytowej, to integracja algorytmów uczenia maszynowego

[1] https://orcid.org/0009-0003-6898-7155, e-mail: krzysztofluczka@outlook.com

pozwala na zwiększenie dokładności i efektywności prognoz. Jednocześnie podkreślono konieczność zapewnienia interpretowalności modeli oraz zachowania równowagi między automatyzacją decyzji a czynnikiem ludzkim.

Słowa kluczowe: ocena zdolności kredytowej, scoring kredytowy, uczenie maszynowe, modele predykcyjne, Biuro Informacji Kredytowej, technologie finansowe, algorytmy klasyfikacyjne

Abstract

The aim of this chapter is to present the evolution of creditworthiness assessment methods and the role of modern technologies in this process. The analysis encompasses both classical statistical approaches and contemporary methods based on machine learning, artificial intelligence, and large language models (LLMs). The first part discusses the historical approach to the problem, from early forms of assessment based on reputation and subjective lender analysis, to the development of the FICO scoring system and the standardisation of credit scoring. This is followed by a presentation of recent solutions, including hybrid models (e.g., TCN-DilateFormer), ensemble algorithms (Random Forest, XGBoost, TabNet), and the application of deep learning methods to credit risk forecasting. The work also discusses the Polish systemic conditions – in particular, the operation of the Polish Credit Information Bureau (BIK) and the importance of "Rekomendacja T" in retail risk management. The conclusions indicate that while traditional methods remain the foundation of credit assessment, integrating machine learning algorithms increases forecast accuracy and efficiency. The need to ensure model interpretability and maintain a balance between decision automation and the human factor is also emphasised.

Keywords: creditworthiness assessment, credit scoring, machine learning, predictive models, Biuro Informacji Kredytowej, financial technologies, classification algorithms

Wstęp

Nowoczesne rozwiązania technologiczne pozwalają nie tylko szybciej gromadzić i przetwarzać ogromne ilości danych, ale również odkrywać wzorce, których człowiek samodzielnie nie byłby w stanie dostrzec. Dzięki technologiom opartym na sztucznej inteligencji, uczeniu maszynowym czy analizie big data możliwe staje się bardziej precyzyjne prognozowanie ryzyka kredytowego i dopasowanie oferty do indywidualnych możliwości klienta. Technologia zmienia

sposób, w jaki instytucje finansowe podchodzą do samej definicji zdolności kredytowej – pozwala ją rozszerzać o nowe źródła danych (takie jak dane cyfrowe, ślady w Internecie, zachowania konsumenckie), a przez to czyni ocenę bardziej kompleksową, dynamiczną i dostosowaną do współczesnych realiów gospodarki cyfrowej.

Zdolność kredytowa to pojęcie bardzo szerokie, które nie ogranicza się jedynie do matematycznego wskaźnika czy prostego porównania przychodów i zobowiązań. W niniejszej pracy będzie oznaczała całościową ocenę możliwości wywiązania się przez klienta z przyszłych zobowiązań finansowych, przy czym obejmuje ona zarówno aspekty finansowe (dochody, wydatki, stabilność źródeł przychodu), jak i elementy behawioralne (historia spłat, wiarygodność, przewidywalność zachowań ekonomicznych), a także szerszy kontekst społeczno-gospodarczy. Jest to wielowymiarowy konstrukt, który łączy w sobie obiektywne dane liczbowe z bardziej jakościowymi informacjami o osobie lub podmiocie ubiegającym się o kredyt.

Rozdział ten stanowi opisowy przegląd literatury Scopus / Web of Science / Google Scholar – nie posiada on ściśle zdefiniowanych kryteriów doboru artykułów, zatem nie jest on ani systematycznym, ani metaanalitycznym opracowaniem całego dorobku badawczego w tym zakresie. Jego celem jest przede wszystkim syntetyczne przedstawienie wybranych publikacji, które w sposób szczególnie wyrazisty ukazują ewolucję metod oceny zdolności kredytowej oraz rolę nowoczesnych technologii w tym procesie. Nacisk został położony na ogólne kierunki rozwoju i główne wnioski badawcze, a nie na szczegółową rekonstrukcję procedur metodologicznych. Za przyjęciem takiej formy stoi fakt, że ze względu na szybkość zmian w obszarze technologii finansowych, przedstawione ustalenia mogą stosunkowo szybko ulec dezaktualizacji. Mimo tych ograniczeń narracyjny charakter przeglądu pozwala uchwycić kluczowe trendy oraz sformułować wnioski istotne z perspektywy dalszych badań nad wpływem technologii na ocenę zdolności kredytowej.

Podstawowe metody oceny zdolności kredytowej

Zdolność kredytowa odzwierciedla wiarygodność finansową jednostki lub przedsiębiorstwa i jest oceniana na podstawie szeregu czynników ilościowych

oraz jakościowych. Współcześnie kluczowe znaczenie mają elementy takie jak historia kredytowa, poziom dochodów, stabilność zatrudnienia, wskaźniki zadłużenia, a także dodatkowe informacje z alternatywnych źródeł danych, np. zachowania konsumenckie, lokalne wskaźniki ekonomiczne czy nawet aktywność w mediach społecznościowych. Ocena ta obecnie wspierana jest zaawansowanymi modelami statystycznymi i uczenia maszynowego. Jednakże wcześniej proces ten miał jednak zupełnie inny charakter – bazował głównie na reputacji klienta, referencjach osobistych, ocenie majątku trwałego (np. ziemi czy nieruchomości) oraz subiektywnym osądzie kredytodawcy.

Badanie Brennecke (2016), oparte na raportach R. G. Dun & Co., ukazuje wczesny etap rozwoju instytucjonalnej oceny historii kredytowej w XIX-wiecznych Stanach Zjednoczonych – przejście od lokalnej, osobistej reputacji kupca do systematycznego, opartego na danych monitorowania wiarygodności finansowej. W latach 1850-1860 informacja kredytowa zaczęła nabierać charakteru zorganizowanego dobra rynkowego: raporty o firmach były regularnie aktualizowane i dostępne dla subskrybentów, co pozwalało tworzyć wczesne profile ryzyka. Jednak, jak pokazuje analiza zachowań użytkowników, informacje te wykorzystywano głównie w reakcji na pogorszenie warunków rynkowych lub negatywne sygnały o dłużnikach, a nie w celu stałej preselekcji. W rezultacie proces oceny kredytowej w omawianym okresie miał charakter reaktywny i kontracykliczny – opierał się na weryfikacji po zdarzeniach, a nie na prewencyjnym screeningu. Ta ewolucja pokazuje, że historia kredytowa w połowie XIX w. była raczej zapisem epizodów niewypłacalności niż ciągłą miarą reputacji finansowej, a dopiero rozwój systematycznego raportowania (i późniejsze ustandaryzowanie punktacji) przekształcił ją w narzędzie predykcyjne używane do oceny ryzyka przed udzieleniem kredytu.

Historyczne badanie Duranda z 1941 roku położyło fundamenty pod nowoczesny credit scoring, wprowadzając po raz pierwszy systematyczne, oparte na danych metody analizy ryzyka kredytowego. Zastąpił subiektywne oceny obiektywną analizą statystyczną, co było rewolucją w branży finansowej lat 40. XX wieku i stało się podstawą dla wszystkich późniejszych rozwojów w dziedzinie oceny ryzyka kredytowego. Autor wskazuje, że kluczowym czynnikiem determinującym ryzyko niewypłacalności jest poziom dochodu kredytobiorcy, przy czym znaczenie mają również takie elementy jak wysokość kredytu, okres spłaty, wkład

własny czy wartość zabezpieczenia. Z czynników niefinansowych największą rolę odgrywają stabilność zatrudnienia i miejsca zamieszkania. Durand opracował narzędzia statystyczne w postaci współczynników efektywności, które pozwalają ocenić wartość informacyjną poszczególnych zmiennych i priorytetyzować je w procesie decyzyjnym. Ważnym wkładem są także proste formuły rankingowe, które mimo swojej prostoty dają wyniki porównywalne ze złożonymi metodami. Zatem Durand pokazuje tym samym, że systematyczna analiza statystyczna przewyższa subiektywną ocenę kredytodawców i może znacząco ograniczyć straty instytucji finansowych, wyznaczając kierunek dla dalszego rozwoju metod ilościowych w ocenie ryzyka kredytowego.

System FICO powstał w drugiej połowie lat 50. XX wieku jako rezultat prac inżynierów i matematyków amerykańskiej firmy Fair, Isaac and Company, którzy przenieśli metody statystyczne i operacyjne, stosowane wcześniej w wojsku i ubezpieczeniach, do sfery masowego kredytowania konsumenckiego (Poon, 2012). Był to moment przełomowy w historii oceny kredytowej, ponieważ po raz pierwszy wprowadzono sformalizowany, ilościowy model punktowy, który pozwalał instytucjom finansowym podejmować decyzje kredytowe w sposób zautomatyzowany, porównywalny i skalowalny. W odróżnieniu od wcześniejszych, subiektywnych metod oceny opartej na reputacji lokalnej, rozmowach z klientem czy intuicji urzędnika, FICO stanowiło próbę standaryzacji zaufania w gospodarce opartej na danych. Jak pokazuje Poon (2012), przełomowość systemu nie wynikała jednak z jego rzekomej obiektywności technologicznej, lecz z faktu, że łączył techniczne modelowanie ryzyka z organizacyjnym uczeniem się instytucji finansowych. Scorecardy FICO były nie tylko narzędziem analitycznym, ale też mechanizmem koordynacji i dyscyplinowania praktyk kredytowych – umożliwiały masowe rozszerzenie rynku konsumenckiego kredytu przy jednoczesnym utrzymaniu kontroli nad ryzykiem.

Z kolei Myers i Forgy (1963) dokonali jednego z pierwszych systematycznych porównań metod ważenia predyktorów w numerycznych systemach oceny ryzyka kredytowego. Analiza została przeprowadzona na próbie 300 wniosków kredytowych (150 „dobrych" i 150 „złych") oraz na odrębnej próbie testowej. Autorzy zestawili skuteczność zarówno metod bardziej zaawansowanych, takich jak klasyczna analiza dyskryminacyjna czy krokowa regresja wieloraka, jak i podejść prostszych, np. przypisania równych wag wszystkim 21 zmiennym. Wyniki wskazały na

porównywalną zdolność rozróżniania kredytobiorców przez obie grupy metod, co dowodzi, że złożone modele nie zawsze przewyższają proste rozwiązania.

W realiach poprzedniego wieku, przy ograniczonych możliwościach obliczeniowych i skromnych zasobach danych, trudno byłoby o lepszą i bardziej systematyczną analizę. To właśnie te badania najlepiej pokazują, jak stopniowo odchodzono od intuicyjnych, subiektywnych ocen kredytodawców w stronę podejścia ilościowego, które stało się standardem w następnych dekadach. Dlatego proponuję traktować je jako punkt wyjścia – zamykający etap wczesnych prób analitycznych i otwierający drzwi do późniejszych badań, opartych już na coraz bardziej zaawansowanych metodach statystycznych i uczenia maszynowego.

Przegląd najnowszych rozwiązań

W ciągu ostatnich czterech dekad obserwujemy wyraźne fale popularności różnych metod analitycznych i uczenia maszynowego. W latach 80. i 90. XX wieku dominowały klasyczne metody statystyczne, takie jak liniowa analiza dyskryminacyjna (LDA), wykorzystywane głównie do klasyfikacji i redukcji wymiarowości danych. W pierwszej dekadzie XXI wieku nastąpił wzrost zainteresowania metodami opartymi na uczeniu maszynowym, takimi jak SVM i Random Forest, które pozwalały na bardziej elastyczne i skuteczne modelowanie złożonych zależności w danych. Jednakże to ostatnie lata, szczególnie od 2020 roku, przyniosły przełom w obszarze dużych modeli językowych (LLM), które rewolucjonizują przetwarzanie języka naturalnego i wspierają szeroki zakres zastosowań od automatyzacji treści po systemy wspomagania decyzji. Te ewolucje metodologiczne znacząco wpływają na współczesne życie, dobrze wpasowując się do systemów finansowych i bankowych. W celu zrozumienia rozwoju i skuteczności różnych metod w kontekście oceny zdolności kredytowej, poniższy przegląd literatury koncentruje się na kluczowych technikach na przestrzeni ostatniej dekady.

W badaniu Shen i Wu (2025) nad predykcją ryzyka kredytowego przedsiębiorstw szczególną uwagę zwraca propozycja hybrydowego modelu TCN-DilateFormer, łączącego zalety sieci konwolucyjnych temporalnych (TCN), które dobrze wyłapują krótkoterminowe wzorce w danych, oraz architektur typu transformer (DilateFormer), które uwzględniają zależności długoterminowe. Autorzy opracowali rozwiązanie ukierunkowane na poprawę dokładności prognozowania

niewypłacalności oraz przezwyciężenie problemu silnie niezbalansowanych danych (stosunek przypadków wypłacalnych do niewypłacalnych wynosił ok. 49:1).

Badanie oparto na danych finansowych 3762 spółek notowanych na rynku chińskim (lata 2012-2022), z czego wykorzystano 132 wskaźniki z sześciu obszarów: wypłacalności, operacyjności, rentowności, rozwoju, struktury wskaźników oraz przepływów pieniężnych. Zastosowano zarówno techniki augmentacji próbek pozytywnych (powielanie, dodawanie szumu Gaussa), jak i downsamplingu próbek negatywnych, co znacząco ograniczyło ryzyko stronniczości modelu. Rezultaty eksperymentalne wskazują, że TCN-DilateFormer przewyższył zarówno tradycyjne metody uczenia maszynowego, jak i inne modele hybrydowe (m.in. CNN-LSTM, CNN-Transformer, ConvTimeNet), osiągając parametry:

- precision = 0.9681 – około 97% firm oznaczonych przez model jako ryzykowne były faktycznie ryzykowne,
- recall = 0.8689 – około 87% faktycznych bankructw zostało wykrytych przez model,
- specificity = 0.9809 – około 98% poprawnie zostało zakwalifikowanych jako bezpieczne.

Test Wilcoxona potwierdził istotność statystyczną wyników, a analiza wrażliwości dodatkowo wykazała stabilność modelu przy zmianach parametrów. W ujęciu praktycznym rozwiązanie to oferuje instytucjom finansowym i inwestorom bardziej precyzyjne narzędzie do oceny ryzyka kredytowego i zarządzania portfelem. Autorzy wskazują na możliwość rozszerzenia podejścia na inne kategorie ryzyka (rynkowe, operacyjne) oraz integrację z dodatkowymi źródłami danych, np. makroekonomicznymi i tekstowymi.

W podobny sposób Xing i in. (2024) przedstawili model zespołowy (ensemble) do przewidywania oceny kredytowej, integrujący trzy różne algorytmy uczenia maszynowego: Random Forest, XGBoost oraz TabNet. Autorzy podjęli problem niezbalansowanych danych w klasyfikacji ryzyka kredytowego, stosując szereg technik przetwarzania, w tym m.in. oversampling oraz metodę SMOTE-ENN – czyli tworzenie nowych sztucznych próbek klasy mniejszościowej (tutaj klasy poor) o cechach pośrednich, jednocześnie czyszcząc zbiór poprzez usunięcie takich punktów, które wyglądają na błędne. Kluczową innowacją było zastosowanie modelowania zespołowego w podejściu stackingowym, w którym Random Forest pełnił rolę meta-modelu integrującego wyniki poszczególnych klasyfikatorów.

Badanie przeprowadzono na zbiorze danych Credit Score Classification (100 000 rekordów, trzy poziomy oceny kredytowej: good, standard, poor). Wyniki empiryczne wykazały, że model zespołowy znacząco przewyższył wydajnością modele pojedyncze, osiągając po zastosowaniu SMOTE-ENN wskaźniki F1-score = 0.7968 (średnia między precision a recall) oraz ROC AUC = 0.9172 (umiejętność modelu w odróżnianiu klasy pozytywnej od negatywnej). Poprawa ta potwierdziła skuteczność zarówno samego podejścia zespołowego, jak i zaawansowanych technik balansowania danych. Autorzy podkreślili, że model zapewnia równowagę między precision i recall, co ma istotne znaczenie w ocenie ryzyka kredytowego, redukując jednocześnie zarówno błędy fałszywie pozytywne, jak i fałszywie negatywne.

Stąpor (2016) przeprowadziła porównanie maszyn wektorów nośnych (z jądrem Gaussa), zestawiając je z heteroskedastyczną analizą dyskryminacyjną (HLDA). Tradycyjna LDA zakłada, że dane w każdej grupie mają rozkład normalny i że wszystkie grupy mają taką samą macierz kowariancji. W praktyce to założenie często jest fałszywe - w grupie osób z problemami finansowymi dane są dużo bardziej „rozproszone" niż w grupie osób stabilnych finansowo, a heteroskedastyczność pozwala to ograniczenie złamać. Badanie oparto na zbiorze danych kredytowych obejmującym 1 000 obserwacji (700 klientów wypłacalnych i 300 niewypłacalnych) oraz 20 cech opisujących profil kredytobiorców. Wyniki wskazały na zbliżoną skuteczność obu podejść: HLDA osiągnęła średnią dokładność 75,10%, podczas gdy SVM uzyskał 76,10%. Różnice te nie były statystycznie istotne, co sugeruje porównywalny potencjał obu metod w ocenie ryzyka kredytowego. Jednocześnie podkreślono, że SVM wiąże się ze znacznie większym kosztem obliczeniowym, ze względu na przykładowo intensywne strojenie hiperparametrów lub większą liczbę wektorów wspierających, podczas gdy HLDA zapewnia prostsze i mniej zasobochłonne rozwiązanie. Autorka rekomenduje HLDA w scenariuszach wymagających efektywności obliczeniowej i prostoty wdrożenia, natomiast SVM jest korzystnym wyborem w warunkach dostępności większych zasobów obliczeniowych i potrzeby maksymalizacji jakości klasyfikacji.

Norling i Abdu (2023) przeanalizowali zastosowanie regresji logistycznej do prognozowania zdolności kredytowej w oparciu o dane ze szwedzkiego instytutu kredytowego, obejmujące około 15 tys. wniosków kredytowych z lat 2020-2022. Modele były budowane dla różnych horyzontów spłaty (6-21 miesięcy),

z wykorzystaniem inżynierii cech (m.in. dyskretyzacja zmiennych ciągłych, selekcja zmiennych wsteczna) oraz kontroli kolinearności. Wyniki pokazały, że regresja logistyczna przewyższa klasyfikator losowy, lecz charakteryzowała się niską czułością wobec klientów niewypłacalnych (AUC ok. 0,290,31). Istotne predyktory zmieniały się w zależności od horyzontu czasowego – krótkoterminowo największe znaczenie miały wcześniejsze zaległości, a długoterminowo informacje o zatwierdzonych kredytach konsumenckich i dochodach.

Z kolei Mestiri (2024) prezentuje systematyczne porównanie tradycyjnych modeli statystycznych, technik uczenia maszynowego oraz metod głębokiego uczenia. Autor uwzględnił sześć podejść: LDA, regresję logistyczną, drzewa decyzyjne, Random Forest (RF), SVM oraz głębokie sieci neuronowe (DNN). Analiza została oparta na próbie 688 obserwacji opisanych dwunastoma zmiennymi finansowymi klientów. Wyniki badania jednoznacznie wskazały przewagę algorytmów uczenia maszynowego i głębokiego uczenia nad klasycznymi technikami statystycznymi. Szczególnie wyróżniły się RF, SVM oraz DNN, które wykazały wyższą zdolność predykcyjną i lepsze dopasowanie do danych.

Bardzo podobne wnioski popiera praca Mbonigaba (2025). Przeprowadził badanie nad skutecznością różnych technik regresji statystycznej w ocenie ryzyka kredytowego w finansach konsumenckich. Autor wykorzystał dane wtórne z lat 2020-2024 i zastosował ilościowy projekt badawczy, porównując klasyczne metody statystyczne z bardziej zaawansowanymi podejściami uczenia maszynowego. W zakresie metodologii uwzględniono regresję logistyczną, regresję liniową wielokrotną, SVM oraz sieci neuronowe. Analiza wykazała, że choć regresja logistyczna pozostaje solidnym narzędziem w tradycyjnym scoringu kredytowym (dokładność 85,3%, AUC = 0,78), to zaawansowane modele osiągnęły istotnie wyższą skuteczność predykcyjną – SVM (92,3%) i sieci neuronowe (91,1%). Badanie pozwoliło także zidentyfikować kluczowe predyktory zdolności kredytowej: długość historii kredytowej (korelacja z niewypłacalnością -0,74), poziom dochodów (-0,52) oraz wiek (-0,26). Ogólny współczynnik korelacji (-0,62) potwierdził silną odwrotną zależność między zmiennymi predykcyjnymi a ryzykiem kredytowym. Wnioski podkreślają, że choć metody tradycyjne zachowują swoją użyteczność, to modele hybrydowe i integracja uczenia maszynowego wyraźnie poprawiają dokładność prognoz. Autor rekomenduje m.in. włączenie alternatywnych źródeł danych kredytowych, regularne aktualizowanie modeli oraz

stosowanie strategii przeciwdziałania stronniczości, aby zwiększyć sprawiedliwość decyzji kredytowych.

Z kolei praca autorstwa Golec i Alabduljalil (2025) stanowi pierwszą próbę kompleksowego ujęcia zastosowań dużych modeli językowych (LLM) w kontekście szacowania ryzyka kredytowego. Nie nadają się do oceny ryzyka kredytowego, ponieważ zostały zaprojektowane do przetwarzania i generowania języka naturalnego, a nie do precyzyjnej analizy danych finansowych i szacowania prawdopodobieństwa niewypłacalności. Jednakże autorzy spojrzeli na problem z perspektywy analizy dokumentów finansowych. Bazując na metodologii PRISMA, dokonali analizy 60 publikacji z lat 2020-2025, co pozwoliło im wyodrębnić cztery główne wymiary klasyfikacji:

- architektury modeli – jaki rodzaj modelu należy zastosować?
- modalności danych – jaki rodzaj danych? Przykładowo: tradycyjne dane jak zarobki czy historia kredytowa, teksty finansowe takie jak raporty czy dokumenty firm oraz dane o zachowaniu klienta w postaci formy korzystania z konta bankowego.
- mechanizmy interpretowalności – jak wytłumaczyć decyzję modelu?
- domena zastosowań – do czego można wykorzystywać LLMy?

W obszarze architektur modeli zidentyfikowano zarówno rozwiązania bazujące na klasycznych enkoderach (np. FinBERT), jak i dekoderach generatywnych (GPT-3.5, GPT-4), a także podejścia hybrydowe i specjalistyczne modele finansowe (FinGPT, FinLLaMA). Analiza modalności danych wskazała na stopniowe odchodzenie od wyłącznie danych strukturalnych na rzecz nieustrukturyzowanych tekstów finansowych czy szeregów czasowych. W zakresie interpretowalności dominują metody post hoc (takie jak SHAP, LIME), jednak rośnie zainteresowanie wyjaśnieniami generowanymi bezpośrednio w procesie wnioskowania oraz projektowaniem modeli wewnętrznie transparentnych. Jeśli chodzi o domeny zastosowań, prace koncentrują się na ocenie zdolności kredytowej, analizie nastrojów rynkowych, a także wykrywaniu oszustw i przeciwdziałaniu praniu pieniędzy. Autorzy wskazują na rosnącą popularność hybryd LLM-ów z klasycznymi metodami uczenia maszynowego oraz dominację metod post hoc. Jednocześnie autorzy identyfikują istotne luki badawcze:

- brak rozwiniętych technik interpretowalności przyczynowej – modele funkcjonują jak „czarne skrzynki", tj. trudno zrozumieć jak dokładnie podejmują decyzje,
- ograniczoną reprodukowalność wyników – dane najczęściej są małe i bardzo zróżnicowane, by móc odtworzyć te same wyniki,
- podatność modeli na stronniczość i halucynacje – niektóre modele mogą być stronnicze wobec pewnych grup społecznych,
- niedobory w zakresie ujednoliconego benchmarkingu – brakuje wspólnych standardów do porównywania wyników oceny zdolności kredytowej.

Figini i Giudici (2017) zaproponowali zastosowanie Bayesowskiego Łączenia Wiedzy (BMA) w ocenie ryzyka kredytowego jako alternatywy dla tradycyjnej regresji logistycznej opartej na pojedynczym modelu. Autorzy wskazali, że w warunkach braku jednoznacznych wskazówek teoretycznych dotyczących doboru zmiennych predykcyjnych, BMA umożliwia formalne uwzględnienie niepewności modelowej poprzez łączenie predykcji wielu konkurencyjnych modeli w średnią ważoną ich prawdopodobieństw posteriori. Wyniki empiryczne potwierdziły, że podejście to zapewnia lepszą kalibrację i stabilność prognoz, a także nieznacznie wyższą wartość AUC w porównaniu z najlepszym pojedynczym modelem regresji logistycznej.

Ocena zdolności kredytowej w Polsce

Biuro Informacji Kredytowej (BIK) zostało utworzone w 1997 roku z inicjatywy Związku Banków Polskich i 21 największych banków w kraju jako odpowiedź na problemy z niewystarczającym przepływem informacji w polskim sektorze bankowym po transformacji ustrojowej. W pierwszej połowie lat 90. brak dostępu do danych o historii kredytowej klientów prowadził do wysokiego udziału należności zagrożonych, sięgającego często ponad 25%, co osłabiało stabilność systemu finansowego. Celem powołania BIK było ograniczenie ryzyka kredytowego, zwiększenie bezpieczeństwa obrotu pieniężnego oraz zapobieganie nadmiernemu zadłużeniu poprzez systematyczne gromadzenie, przetwarzanie i udostępnianie informacji o zobowiązaniach kredytowych klientów banków i instytucji finansowych. Jak podkreśla Związek Banków Polskich (2013), utworzenie BIK umożliwiło wprowadzenie nowoczesnych mechanizmów oceny

zdolności kredytowej i znacząco poprawiło efektywność oraz bezpieczeństwo krajowego rynku kredytowego.

Według danych Biura Informacji Kredytowej (2019) średni wynik punktowy BIK w Polsce na koniec czerwca 2019 roku wynosił 79 punktów, co wskazuje na relatywnie wysoki poziom wiarygodności kredytowej wśród polskich konsumentów. Ponad połowa kredytobiorców (ok. 7 mln osób) uzyskała ocenę w najwyższym przedziale 80-100 punktów, co oznacza niskie ryzyko kredytowe i wysokie prawdopodobieństwo terminowej spłaty zobowiązań. Wynik BIK opiera się na czterech podstawowych komponentach: terminowości spłat, stopniu wykorzystania limitów kredytowych, liczbie zapytań o kredyt oraz długości historii kredytowej (wraz ze zgodą na przetwarzanie danych). Analizy przestrzenne wskazują na koncentrację najwyższych ocen w dużych miastach oraz w regionach o stabilnej strukturze gospodarczej, m.in. na Podkarpaciu (Rzeszów – 82,2; Krosno – 81,9). Wskaźnik scoringowy można było obliczyć dla ok. 86% aktywnych klientów (13,4 mln osób), co potwierdza wysoki poziom upowszechnienia systemu oceny kredytowej w Polsce. Dane te dowodzą, że scoring BIK stał się kluczowym narzędziem oceny ryzyka kredytowego i elementem infrastruktury finansowej kraju.

Ważnym elementem do poruszenia jest również Rekomendacja T stanowi ona dokument regulacyjny opracowany przez Komisję Nadzoru Finansowego (KNF), którego celem jest ujednolicenie i usystematyzowanie zasad zarządzania ryzykiem detalicznych ekspozycji kredytowych w polskim sektorze bankowym. Została wprowadzona w lutym 2010 roku jako reakcja na globalny kryzys finansowy i niewystarczające zabezpieczenia systemów oceny ryzyka kredytowego wynikające z wcześniejszych ram Bazylea II. Jej aktualna wersja (KNF, 2013) określa dobre praktyki w zakresie identyfikacji, pomiaru, monitorowania i kontroli ryzyka kredytowego dla banków działających w Polsce, ze szczególnym uwzględnieniem kredytów konsumenckich i innych detalicznych form zadłużenia.

Zgodnie z dokumentem KNF (2013), Rekomendacja T nakłada na banki obowiązek prowadzenia pisemnej polityki ryzyka detalicznego, określania apetytu na ryzyko oraz limitu wskaźnika długu do dochodu. Wymaga również rozdziału funkcji sprzedażowych od analitycznych, wprowadzenia systematycznego monitoringu portfela kredytowego, testów warunków skrajnych, a także korzystania z zewnętrznych baz danych, takich jak Biuro Informacji Kredytowej (BIK)

i biura informacji gospodarczej. Dokument promuje zasadę konserwatywnej oceny zdolności kredytowej, przejrzystych relacji z klientem oraz proporcjonalności procedur do skali działalności banku. Jak zauważa Strycharczyk (2012), wprowadzenie Rekomendacji T było odpowiedzią na obserwowane w latach poprzednich zbyt liberalne praktyki kredytowe, takie jak wydłużanie okresów spłaty, zaniżanie wymogów dotyczących zabezpieczeń czy udzielanie kredytów osobom o niskiej zdolności kredytowej. Autorka podkreśla, że wdrożenie Rekomendacji przyniosło istotne korzyści w postaci poprawy jakości portfeli kredytowych, wzrostu przejrzystości procesów decyzyjnych i ochrony konsumentów przed nadmiernym zadłużeniem, ale jednocześnie wiązało się z pewnymi zagrożeniami. Do najczęściej wskazywanych należały: wzrost kosztów wdrożenia nowych procedur, ograniczony dostęp do kredytu dla gospodarstw o niższych dochodach oraz ryzyko przenoszenia części popytu na kredyty do sektora parabankowego.

W efekcie zarówno Biuro Informacji Kredytowej, jak i Rekomendacja T stanowią kluczowe narzędzia wspierające odpowiedzialne udzielanie kredytów, ograniczanie ryzyka finansowego i zwiększanie stabilności polskiego sektora bankowego. Z kolei w polskiej literaturze naukowej zagadnienie oceny zdolności kredytowej rozwijało się równolegle z trendami światowymi, choć często w większym stopniu uwzględniało lokalne uwarunkowania gospodarcze i instytucjonalne. Zatem w tym rozdziale zostaną omówione wybrane polskie prace naukowe, które pokazują, w jaki sposób definiuje się i bada zdolność kredytową w krajowym kontekście oraz jakie rozwiązania są proponowane w praktyce.

W pracy Juszczyk (2008) podjęto próbę systematycznego ujęcia oceny zdolności kredytowej małych i średnich przedsiębiorstw w praktyce banków komercyjnych i spółdzielczych w Polsce. Badanie oparto na wywiadach z kadrą kierowniczą i merytoryczną banków oraz analizie literatury i uregulowań prawnych. Autorzy wskazali, że zdolność kredytowa obejmuje zarówno aspekt formalno-prawny, jak i merytoryczny, przy czym ten drugi dzieli się na czynniki personalne (zaufanie do kredytobiorcy) i ekonomiczne (takie jak sytuacja finansowa, zabezpieczenia). Szczególną uwagę poświęcono różnorodnym metodom oceny: opisowej (listy kontrolne i klasy ryzyka), punktowej (system wag i oceny wskaźników), finansowej (analiza rentowności, płynności, zadłużenia i sprawności działania), dyskryminacyjnej (m.in. model Altmana), a także podejściom jakościowym, jak CAMPARI ICE czy 5C/6C. Wskazano również na znaczenie oceny

projektów inwestycyjnych z wykorzystaniem NPV, IRR czy progów rentowności. Wyniki podkreślają, że mali i średni przedsiębiorcy stanowią niejednorodną grupę kredytobiorców, co wymaga elastycznego doboru metod – od prostszych analiz w przypadku mniejszych kredytów po pełne procedury wskaźnikowe przy większych przedsięwzięciach. Autorzy rekomendują łączenie metod ilościowych i jakościowych, dostosowanie analizy do specyfiki branżowej oraz regularną aktualizację modeli, co stanowi fundament współczesnych praktyk oceny ryzyka kredytowego w Polsce.

Istotny wkład w dyskusję nad samą problematyką oceny zdolności kredytowej wniosła praca Nowaka (2017), której celem była identyfikacja głównych nieprawidłowości w tym procesie w odniesieniu do klientów detalicznych. Autor, opierając się na analizie raportów organów nadzoru, literatury oraz regulacji prawnych, wskazał na liczne uchybienia zarówno po stronie banków, jak i kredytobiorców. Po stronie instytucji finansowych najczęściej powtarzającymi się problemami były m.in. niedoszacowanie ryzyka stopy procentowej, zaniżanie kosztów utrzymania gospodarstw domowych, nadmierne wydłużanie okresu kredytowania czy zbyt liberalne zasady przyznawania kredytów. Z kolei konsumenci nierzadko wykazywali brak rzetelności (np. poprzez zatajenie istotnych informacji) oraz deficyty kompetencji finansowych, czego przejawem była m.in. niechęć do czytania umów kredytowych przed ich podpisaniem. Autor zwrócił także uwagę na konieczność działań naprawczych, zarówno po stronie regulatorów (np. rekomendacje KNF i EBA), jak i rynku, w tym edukacji finansowej konsumentów. Wnioski tej pracy ukazują, że proces oceny zdolności kredytowej ma charakter systemowy i wymaga kompleksowych działań, obejmujących jednocześnie banki, konsumentów oraz organy nadzoru.

W literaturze polskiej na uwagę zasługuje praca Czech (2021), która podjęła próbę oceny ryzyka kredytowego Polski w okresie pandemii COVID-19 na podstawie ubezpieczenia SCDS na wypadek niewypłacalności państwa. Badanie objęło okres od marca 2019 do lutego 2021 i wykorzystało zarówno analizę dynamiki, jak i modele regresji liniowej z uwzględnieniem czynników makroekonomicznych oraz nieekonomicznych. Wyniki wskazały, że w początkowej fazie pandemii spready SCDS wzrosły, jednak z czasem powróciły do trendu spadkowego, mimo wzrostu długu publicznego i pogorszenia wskaźników gospodarczych. Szczególnie interesujące jest odkrycie, iż determinanty ryzyka

kredytowego różniły się przed i w trakcie pandemii – w okresach stabilnych dominowały zmienne makroekonomiczne, natomiast w czasie kryzysu istotną rolę odgrywały również czynniki związane z przebiegiem pandemii, takie jak liczba zgonów czy zakażeń. Autorka dochodzi do wniosku, że spready SCDS dobrze odzwierciedlają poziom ryzyka kredytowego w okresach równowagi, jednak ich użyteczność w warunkach niepewności generowanej przez czynniki pozaekonomiczne jest ograniczona.

Polski rynek kredytowy i jego badanie ukazują złożoność zagadnienia oceny zdolności kredytowej zarówno na poziomie mikro, jak i makro. Badacze podkreślają, że ocena ta wymaga podejścia wielowymiarowego, łączącego analizę ilościową i jakościową, a także uwzględniającego czynniki instytucjonalne i koniunkturalne. Z jednej strony wskazuje się na potrzebę stosowania różnorodnych metod i dostosowania ich do specyfiki kredytobiorcy, z drugiej na ryzyka wynikające z nieprawidłowości proceduralnych czy asymetrii informacji. Jednocześnie najnowsze badania dowodzą, że tradycyjne narzędzia oceny ryzyka kredytowego mogą tracić skuteczność w warunkach kryzysów o charakterze pozaekonomicznym, co potwierdza konieczność ciągłej aktualizacji i poszerzania metodologii.

Podsumowanie i wnioski

Według przeglądu literatury Markova i in. (2022) literatura nie dostarcza jednoznacznej odpowiedzi na to, która metoda jest najbardziej efektywna. Wykazano natomiast, że w liczbie zastosowań dominuje głównie regresja logistyczna, a dopiero na kolejnych miejscach znajdują się SVM i sieci neuronowe. Z niniejszego przeglądu literatury również nie można jednoznacznie określić, która metoda oceny zdolności kredytowej jest najskuteczniejsza, choć z analizy przytoczonych badań wynika, że regresja logistyczna, mimo iż pozostaje solidnym narzędziem bazowym w credit scoringu, jest generalnie mniej skuteczna niż zaawansowane metody uczenia maszynowego i głębokiego uczenia.

Mimo dynamicznego rozwoju metod ilościowych i technologii AI, wciąż istnieje wiele obszarów wymagających pogłębionych badań. Po pierwsze, konieczne jest lepsze zrozumienie wpływu nietypowych i nieustrukturyzowanych źródeł danych – takich, jak aktywność w mediach społecznościowych czy dane behawioralne z urządzeń mobilnych – na trafność i uczciwość oceny kredytowej,

zwłaszcza w kontekście prywatności i ochrony danych. Po drugie, brakuje spójnych ram do oceny ryzyka algorytmicznego i modelowego, w tym standardów interpretowalności oraz mechanizmów wykrywania i korygowania stronniczości czy „halucynacji" w modelach uczenia maszynowego. Trzecim niewyjaśnionym obszarem pozostaje integracja metod predykcyjnych z systemami decyzyjnymi banków w sposób, który zapewni zarówno zgodność z regulacjami (np. Rekomendacja T, Basel III), jak i elastyczność operacyjną. Wreszcie, wymaga się badań nad skutecznością hybrydowych modeli człowiek-maszyna, które określą, jak najlepiej łączyć intuicyjne, jakościowe oceny ekspertów z precyzyjnymi, ale złożonymi algorytmami, aby maksymalizować zarówno trafność decyzji kredytowych, jak i ochronę konsumentów.

W Polsce ocena zdolności kredytowej wciąż pozostaje hybrydą metod tradycyjnych i nowoczesnych. Z jednej strony mocno zakorzenione są rozwiązania regulacyjne i jakościowa ocena klienta, z drugiej strony coraz wyraźniej widać wpływ technologii i modeli ilościowych. W artykule podkreślono, że największym wyzwaniem nie jest dziś brak narzędzi, ale odpowiedzialne ich stosowanie oraz równoległa edukacja finansowa społeczeństwa, bo tylko wtedy ocena zdolności kredytowej będzie rzetelna i wiarygodna. Należy zaznaczyć, że nie można całkowicie wyeliminować czynnika ludzkiego – doświadczenie, intuicja i możliwość uwzględnienia kontekstu poza surowymi danymi nadal są nieocenione. Zaleca się, zatem model hybrydowy człowiek-maszyna, w którym system podejmuje decyzję kredytową wyłącznie w sytuacjach jednoznacznych, gdy wszystkie kluczowe wskaźniki wskazują na ten sam kierunek. W przypadkach granicznych lub niespójnych ostateczne rozstrzygnięcie pozostaje w gestii człowieka.

Przytoczone argumenty wskazują jednoznacznie, że ogólny wpływ technologii na ocenę zdolności kredytowej jest zdecydowanie pozytywny. Umożliwia szybsze, bardziej kompleksowe i dokładniejsze prognozowanie ryzyka, a także włączenie nowych źródeł danych do analizy. Jednocześnie zwrócono uwagę, że nadal pozostaje ogromne pole do rozwoju, szczególnie w zakresie przejrzystości modeli i lepszej interpretowalności wyników. Powinniśmy dążyć do systemu, w którym kluczowe znaczenie zachowa historia kredytowa i dane twarde, ponieważ to one w największym stopniu odzwierciedlają rzeczywistą wiarygodność finansową jednostki.

Literatura

1. Brennecke, C. (2016). Information Acquisition in Antebellum U.S. Credit Markets: Evidence from Nineteenth-Century Credit Reports. *FDIC Center for Financial Research Working Paper No. 2016-04.*
2. Czech, M. (2021). Assessment of the credit risk of Poland based on sovereign credit default swap spreads during the COVID-19 pandemic. *Ekonomia i Prawo. Economics and Law, 20*(3), 497-511. https://doi.org/10.12775/EiP.2021.030
3. Durand, D. (1941). *Risk elements in consumer instalment financing*. New York: National Bureau of Economic Research.
4. Figini, S., Giudici, P. (2017). Credit risk assessment with Bayesian model averaging. *Communications in Statistics – Theory and Methods, 46*(19), 9507-9517. https://doi.org/10.1080/03610926.2016.1212070
5. Golec, M., AlabdulJalil, M. (2025). Interpretable LLMs for credit risk: A systematic review and taxonomy (arXiv:2506.04290v2) [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2506.04290
6. Juszczyk, S. (2008). Zdolność kredytowa w ocenie banku na przykładzie małych i średnich przedsiębiorstw. *Ekonomiczne i Organizacyjne Problemy Bankowości, 66*, 17-27. Pobrane z: https://sj.wne.sggw.pl/pdf/EIOGZ_2008_n66_s17.pdf
7. Markov, A., Seleznyova, Z., Lapshin, V. (2022). Credit scoring methods: Latest trends and points to consider. *The Journal of Finance and Data Science, 8,* 180-201. https://doi.org/10.1016/j.jfds.2022.07.002
8. Mbonigaba, C. (2025). Evaluating credit scoring models with statistical regression techniques in consumer finance (Version March 01, 2025). *SSRN*. https://doi.org/10.2139/ssrn.5188111
9. Mestiri, S. (2024). Credit scoring using machine learning and deep learning-based models. *Data Science in Finance and Economics, 4*(2), 236-248. https://doi.org/10.3934/DSFE.2024009
10. Myers, J. H., Forgy, E. W. (1963). The development of numerical credit evaluation systems. *Journal of the American Statistical Association, 58*(303), 799-806. https://doi.org/10.2307/2282727
11. Nowak, P. (2017). The incorrectness in the assessment procedure of creditworthiness. *Journal of Finance and Financial Law, 2*(14), 53-65. https://doi.org/10.18778/2391-6478.2.14.05

12. Shen, C., Wu, J. (2025). Research on credit risk of listed companies: A hybrid model based on TCN and DilateFormer. *Scientific Reports, 15*, 2599. https://doi.org/10.1038/s41598-025-86371-7
13. Stąpor, K. (2016). *A critical comparison of discriminant analysis and SVM-based approaches to credit scoring*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach. Pobrane z: http://sbc.org.pl/Content/264089/PDF/05.pdf
14. Strycharczyk, A. (2012). Rekomendacja T – szanse i zagrożenia. *Problemy Zarządzania, Finansów i Marketingu, 27*, 545-554. Pobrane z: https://bazhum.muzhp.pl/media/texts/problemy-zarzadzania-finansow-i-marketingu/2012-tom-27/problemy_zarzadzania_finansow_i_marketingu-r2012-t27-s545-554.pdf
15. Xing, Q., Yu, C., Huang, S., Zheng, Q., Mu, X., Sun, M. (2024). Enhanced credit score prediction using an ensemble deep learning model. *arXiv*. https://doi.org/10.48550/arXiv.2410.00256

Netografia

1. Biuro Informacji Kredytowej. (2019). *BIK: Prawie 7 mln Polaków ma doskonały scoring*. Informacja prasowa. Pobrane z: https://media.bik.pl/informacje-prasowe/454131/bik-prawie-7-mln-polakow-ma-doskonaly-scoring (dostęp: 01.09.2025).
2. Komisja Nadzoru Finansowego. (2013). *Rekomendacja T dotycząca dobrych praktyk w zakresie zarządzania ryzykiem detalicznych ekspozycji kredytowych*. Pobrane z: https://www.knf.gov.pl/knf/pl/komponenty/img/rekomendacja_t_%2814_09_2018%29_63160.pdf (dostęp: 01.09.2025).
3. Norling, J., Abdu, S. (2023). From data to decision: Using logistic regression to determine creditworthiness (Bachelor's thesis, KTH Royal Institute of Technology). *KTH Royal Institute of Technology.* Pobrane z: https://kth.diva-portal.org/smash/record.jsf?pid=diva2:1833653 (dostęp: 01.09.2025).
4. Poon, M. A. (2012). What lenders see – A history of the Fair Isaac scorecard [Doctoral dissertation, University of California, San Diego]. *eScholarship.* Pobrane z: https://escholarship.org/uc/item/7n1369x2 (dostęp: 01.09.2025).
5. Związek Banków Polskich. (2013). *Jak powstało BIK*. Biuro Informacji Kredytowej. Pobrane z: https://www.skef.pl/wp-content/uploads/2013/07/Jak-powstało-BIK.pdf (dostęp: 01.09.2025).