



www.aileaders-project.eu

Stronnicze modele sztucznej inteligencji wykorzystywane do oceny zdolności kredytowej



Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

STUDIUM PRZYPADKU: 01

Niniejszy OER stanowi studium przypadku dotyczące tego, w jaki sposób stronnicze lub niereprezentatywne dane mogą prowadzić do niepożądanych skutków, dyskryminacji i wykluczenia poprzez wykorzystanie automatycznych usług oceny zdolności kredytowej. Studium przypadku analizuje pojawienie się i rozwój automatycznych systemów podejmowania decyzji kredytowych, które wykorzystują algorytmy do analizy różnych punktów danych w celu oceny zdolności kredytowej danej osoby i podjęcia decyzji o zatwierdzeniu lub odrzuceniu wniosku kredytowego.

Studium przypadku koncentruje się na tym, jak działają systemy oceny zdolności kredytowej, wykorzystując symulację do zilustrowania, jakie punkty danych gromadzą i jakie rodzaje danych osobowych wykorzystują do podejmowania decyzji. Studium przypadku zwiększa również świadomość na temat problemów generowanych przez stronnicze lub niereprezentatywne zbiory danych oraz tego, w jaki sposób mogą one sprawić, że systemy oceny zdolności kredytowej będą niesprawiedliwe, funkcjonując jako bariery, a nie ulepszenia w zakresie dostępu obywateli do usług kredytowych i pożyczkowych na sprawiedliwych warunkach.

Cel/przeznaczenie

Celem tego studium przypadku jest podniesienie świadomości na temat potencjału automatycznych systemów oceny zdolności kredytowej w zakresie poprawy dostępu do kredytów, a także wyzwań i zagrożeń dla prywatności i dyskryminacji, jakie wiążą się z wykorzystaniem tych systemów, zwłaszcza w przypadku stosowania stronniczych zbiorów danych, a także gdy modele są szkolone na danych, które nie są reprezentatywne dla niektórych populacji.



Oczekiwane efekty kształcenia

- 01** Student będzie potrafił **zidentyfikować** ryzyko etyczne w modelach oceny zdolności kredytowej i zaproponować środki zaradcze.
- 02** Student **zrozumie** kluczowe kwestie etyczne związane z zastosowaniem sztucznej inteligencji w udzielaniu kredytów i ocenie zdolności kredytowej, w tym stronniczość, dyskryminację, przejrzystość i prywatność danych.

Sugerowane podejście metodologiczne

Nauczyciele powinni zachęcać uczniów do samodzielnego myślenia i identyfikowania innych potencjalnych problemów, które mogą mieć. Przykłady i dodatkowe lektury są dostępne za pośrednictwem linków, a także dostęp do symulacji mającej na celu Przedstawiono tematy do dyskusji, obawy i potencjalne rozwiązania, ale wykładowcy powinni zachęcać studentów do samodzielnego myślenia i identyfikowania innych potencjalnych problemów, które mogą ich niepokoić. Przykłady i dodatkowe materiały do przeczytania są dostępne za pośrednictwem linków, a także poprzez dostęp do symulacji mającej na celu zilustrowanie, jakie dane osobowe są wykorzystywane w automatycznej ocenie zdolności kredytowej i jakie wyniki dają te systemy.



Słowa kluczowe

Udzielanie kredytów, ocena zdolności kredytowej, ochrona prywatności i danych osobowych, przejrzystość, stronniczość, dyskryminacja

Automatyczny dostęp do kredytu

Być może widzieliście Państwo w Internecie oferty firm wydających karty kredytowe, które obiecują zatwierdzenie nowej karty kredytowej (i odpowiedniego limitu kredytowego) w ciągu zaledwie jednego lub dwóch dni. Niektóre z nich obiecują obliczenie limitu kredytowego w ciągu kilku minut i natychmiastowe zezwolenie na jego wykorzystanie.

Być może widzieliście Państwo w Internecie oferty firm wydających karty kredytowe, które obiecują zatwierdzenie zupełnie nowej karty kredytowej (i odpowiedniego limitu kredytowego) w ciągu zaledwie jednego lub dwóch dni. Niektóre z nich obiecują obliczenie limitu kredytowego w ciągu kilku minut i natychmiastowe zezwolenie na jego wykorzystanie. Automatyczne wnioski kredytowe online ułatwiły i znacznie przyspieszyły ten proces, umożliwiając niemal natychmiastowe obliczenie zdolności kredytowej. Dzięki temu firmy mogą uzyskać automatyczne rekomendacje i/lub zatwierdzenia dotyczące szeregu produktów finansowych – nie tylko kart kredytowych, ale także innych produktów, takich jak kredyty osobiste lub hipoteczne, a jednocześnie obliczyć i przedstawić ofertę dotyczącą oprocentowania, jakie należy zapłacić za każdy z tych produktów.

Wyniki kredytowe to modele komputerowe, które korelują szereg czynników z prawdopodobieństwem niewywiązania się z płatności zadłużenia. Innymi słowy, wynik kredytowy odzwierciedla zdolność kredytową danej osoby lub jej zdolność i wiarygodność w zakresie spłaty danej pożyczki. Wyniki kredytowe opierają się na historii kredytowej danej osoby, zbudowanej w oparciu o szereg danych, takich jak liczba posiadanych rachunków i czas ich istnienia, całkowity poziom zadłużenia, historia spłat, rodzaje posiadanych kredytów, długość transakcji kredytowych, proporcja wykorzystanego zadłużenia oraz to, czy w danym okresie wnioskowano o nowe rachunki. Wykorzystując te dane, oceny zdolności kredytowej pozwalają oddzielić dobre ryzyko kredytowe od złego i sklasyfikować potencjalnych kredytobiorców w celu przewidzenia prawdopodobieństwa niewywiązania się z zobowiązań.

W sektorze finansowym banki i instytucje finansowe wykorzystują oceny kredytowe do podejmowania decyzji dotyczących kredytów lub kart kredytowych; firmy ubezpieczeniowe wykorzystują je do oceny profilu ryzyka ubezpieczonych; sprzedawcy detaliczni wykorzystują je do planów ratalnych; salony samochodowe wykorzystują je do oceny kwalifikowalności i warunków finansowania samochodów; a kredytodawcy hipoteczni wykorzystują je do oceny zdolności kredytowej nabywców nieruchomości. Nawet operatorzy telefoniczni wykorzystują je do podejmowania decyzji o tym, czy dana osoba kwalifikuje się do planów przedpłaconych lub abonamentowych, lub do ustalenia, czy mogą zaoferować jej nowy telefon finansowany i spłacany w ratach. Połączenie wielu punktów danych i wielowymiarowych założeń jest statystycznie łączone i ważne, aby umożliwić automatyczne podejmowanie decyzji kredytowych.

Sztuczna inteligencja i uczenie maszynowe wykorzystują modele statystyczne i analizę danych, aby przyspieszyć i usprawnić proces oceny zdolności kredytowej. Pozwala to na większą szybkość i potencjalnie większą dokładność oraz pewność decyzji kredytowych. W ciągu ostatnich kilku lat zarówno sektor bankowy, jak i kredytowy zaczęły oferować kilka produktów **FinTech** (*technologii finansowej*), które zapewniają zautomatyzowane rozwiązania w zakresie podejmowania decyzji kredytowych.

Decyzje kredytowe **oparte na sztucznej inteligencji** mogą potencjalnie **poprawić wydajność i wyniki, obniżając koszty dla instytucji finansowych, co może również przynieść korzyści konsumentom poprzez rozszerzenie dostępu do kredytów lub obniżenie ich kosztów**.¹ „Nowe, wysokowydajne modele pozwalają bankom na bardziej precyzyjne definiowanie parametrów kredytowych (i kapitałowych), a tym samym zwiększają ich zdolność do zatwierdzania wniosków klientów posiadających zdolność kredytową i odrzucania wniosków klientów, którzy nie posiadają zdolności kredytowej lub nie są w stanie spłacić dalszego zadłużenia. W rzeczywistości banki (i firmy z branży fintech), które wdrożyły takie nowe modele, już zwiększyły przychody, zmniejszyły wskaźniki strat kredytowych i osiągnęły znaczną poprawę wydajności dzięki bardziej precyzyjnemu i zautomatyzowanemu podejmowaniu decyzji².” Istnieje jednak kilka zagrożeń i wad, które omówimy w niniejszym studium przypadku. Jedną z nich jest brak wyjaśnienia lub niemożność wyjaśnienia, dlaczego programy podejmują określone decyzje. Kolejnym jest nierówny wpływ lub dyskryminacja, które mogą dotyczyć niektóre grupy bardziej niż inne ze względu na praktyki biznesowe (takie jak sposób podejmowania decyzji kredytowych). Jak zobaczymy, jest to bezpośrednio związane z wykorzystaniem błędnych lub stronniczych danych do podejmowania decyzji kredytowych, co może utrudniać, a nie ułatwiać dostęp do kredytów niektórym grupom społecznym, pogłębiając w ten sposób nierówności w dostępie do kredytów³.

- 1 Congressional Research Service. (2023). *Generative artificial intelligence: Overview, issues, and questions for Congress* (CRS Report No. IF12399). <https://crsreports.congress.gov/product/pdf/IF/IF12399>
- 2 Dash, R., Kremer, A. i Petrov, A. (2021). *Projektowanie modeli podejmowania decyzji kredytowych nowej generacji*. McKinsey & Company. <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models>
- 3 Andrews, E. L. (6 sierpnia 2021 r.). *Jak błędne dane pogłębiają nierówności w zakresie kredytów*. Stanford HAI. <https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>

Modele danych i decyzje kredytowe oparte na sztucznej inteligencji

Kredyty oparte na sztucznej inteligencji – konkretne zalety

Decyzje kredytowe podejmowane przy pomocy sztucznej inteligencji mają potencjalne zalety, które wykraczają poza poprawę wydajności i obniżenie kosztów związanych z udostępnianiem kredytów, dzięki czemu stają się one tańsze.

Potencjalnie **uczenie maszynowe** w ocenie zdolności kredytowej może również zwiększyć liczbę i charakter punktów danych, które mają wpływ na ocenę zdolności kredytowej osób, a także może zaoferować możliwość tworzenia ocen dla społeczności, które od dawna mają ograniczony dostęp do usług bankowych, oraz zapewnić **dostęp do rozwiązań bankowych i kredytowych dla całych społeczności i osób**, które tradycyjnie były wykluczone z formalnych instytucji finansowych. Oferując alternatywne rozwiązania poprzez uwzględnienie szerszego zakresu danych, **ocena oparta na sztucznej inteligencji** „umożliwia ocenę osób fizycznych bez tradycyjnej historii kredytowej poprzez analizę alternatywnych źródeł danych, takich jak transakcje internetowe, interakcje w mediach

społecznościowych, nawyki przeglądania stron internetowych lub korzystanie z aplikacji mobilnych”.⁴ W przypadku osób posiadających tradycyjne dane bankowe i finansowe, ocena oparta na sztucznej inteligencji może zapewnić bardziej precyzyjną i ulepszoną analizę, dzięki czemu decyzje dotyczące oceny kredytowej i udzielania kredytów będą dokładniejsze. Osobom, które nie posiadają tradycyjnej historii kredytowej, może ona zapewnić dostęp do kredytów i innych usług bankowych.

Tradycyjne źródła danych dotyczących oceny zdolności kredytowej

- otwarte karty kredytowe i linie kredytowe
- kredyty samochodowe
- kredyty hipoteczne
- historia spłat kredytów
- historia zapytań kredytowych
- zgłoszenia upadłości

Ilustracja 1. Przykłady tradycyjnych i alternatywnych danych kredytowych.

Źródło: <https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf>

Alternatywne źródła danych dotyczących oceny zdolności kredytowej

- dane dotyczące przepływów pieniężnych
- płatności rachunków
- dane dotyczące wynajmu
- dokumentacja zatrudnienia
- dokumentacja dotycząca wykroczeń drogowych lub sporów
- wzorce użytkowania danych telekomunikacyjnych i mobilnych
- dane z mediów społecznościowych
- dane dotyczące zachowań
- płatności podatków

Obecnie znaczna część światowej populacji jest uznawana za konsumentów niewidocznych kredytowo (bez dostępu) lub o słabej zdolności kredytowej (słaby dostęp w złych warunkach). Szacuje się, że w samych Stanach Zjednoczonych 45 milionów konsumentów nie ma dostępu do tradycyjnych modeli kredytowych lub ma do nich ograniczony dostęp⁵. W krajach takich jak Indie i RPA ponad połowa ludności nie ma dobrych możliwości uzyskania kredytu⁶. Tam, gdzie dostęp był możliwy, istnieje potencjał, aby uczynić go bardziej sprawiedliwym i wydajnym, natomiast tam, gdzie dostęp nie był możliwy, istnieje potencjał, aby w końcu go zapewnić, gwarantując jednocześnie, że będzie on równie wydajny i sprawiedliwy.

Istnieje ogromny potencjał dla bardziej integracyjnych i zrównoważonych modeli oceny zdolności kredytowej, które z kolei mogą sprzyjać postępowi i bezpieczeństwu finansowemu szerszej grupy ludności; więcej osób niż kiedykolwiek wcześniej ma potencjał dostępu do usług bankowych i finansowych, co z kolei może doprowadzić do pojawienia się na rynku bardziej konkurencyjnych produktów kredytowych, które ostatecznie mogą sprawić, że kredyty staną się bardziej przystępne⁷.

Dlaczego więc obawiamy się, że decyzje kredytowe oparte na sztucznej inteligencji mogą zwiększyć, a nie zmniejszyć wykluczenie finansowe?

4 Bag, S. (2024). Sztuczna inteligencja a ocena zdolności kredytowej: zalety algorytmów i środki ostrożności. Observer Research Foundation. <https://www.orfonline.org/expert-speak/ai-and-credit-scoring-the-algorithmic-advantage-and-precaution>

5 TransUnion. (2022). Ponad 45 milionów Amerykanów nie ma dostępu do kredytów lub ma do nich ograniczony dostęp; około 20% z nich co dwa lata staje się aktywnymi kredytobiorcami. TransUnion. https://newsroom.transunion.com/more-than-45-million-americans-are-either-credit-unserved-or-underserved--approximately-20-migrate-to-being-credit-active-every-two-years#_edn1

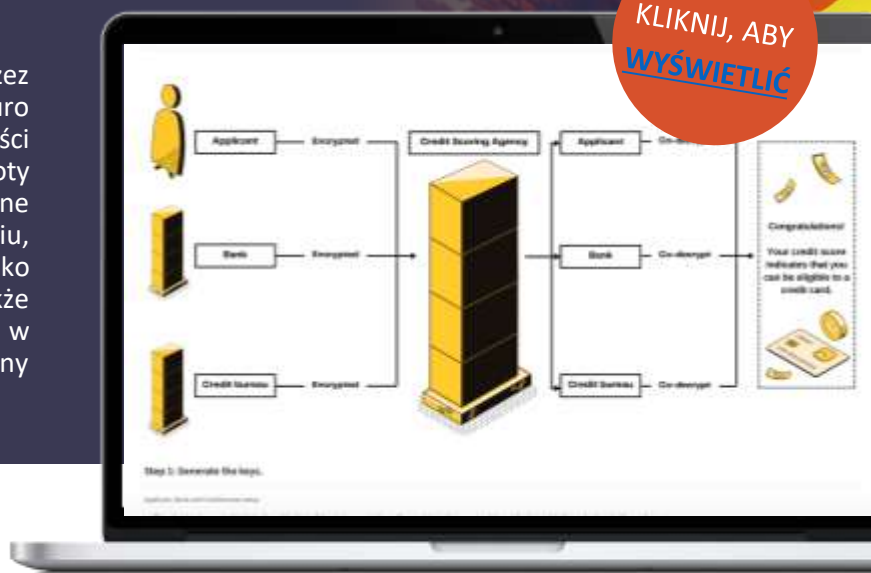
6 Zespół Svitla. (2024). Uczenie maszynowe w ocenie zdolności kredytowej: korzyści, modele i wyzwania związane z wdrożeniem. Svitla Systems. <https://svitla.com/blog/machine-learning-for-credit-scoring/>

7 Alliance for Financial Inclusion. (2025). Alternatywne dane do oceny zdolności kredytowej. <https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf>

Ocena zdolności kredytowej i wykorzystanie danych osobowych

Proces zatwierdzania kredytu – ocena zdolności kredytowej i zatwierdzanie kredytu to złożone procesy, które odbywają się za kulisami, aby umożliwić podjęcie decyzji.

Obejmuje to dane wprowadzone przez wnioskodawcę, bank lub instytucję finansową, biuro informacji kredytowej oraz agencję oceny zdolności kredytowej. Osoba fizyczna i inne podmioty dostarczają do procesu bogate dane, w tym dane osobowe. Poniższy obrazek pokazuje w przybliżeniu, jak działa ten proces. Jak widać, nie tylko przedstawia on zaangażowane podmioty, ale także pokazuje poziomy bezpieczeństwa stosowane w celu ochrony danych wykorzystywanych do oceny zdolności kredytowej i podjęcia decyzji.



Ilustracja 2. Schemat przedstawiający proces oceny zdolności kredytowej i podejmowania decyzji kredytowej.

Źródło: https://huggingface.co/spaces/zama-fhe/encrypted_credit_scoring

The screenshot shows a credit assessment form. The form asks "Which of the following do you actively hold or own?" with options Car, Property (selected), and Mobile phone. It also asks for "Number of children", "Household size", "Income", and "Age". A red circular button with the text "KLIKNIJ, ABY WYŚWIETLIĆ" is overlaid on the form.

Aby zilustrować, jak działa ten proces i jakie dane należy wprowadzić, aby ocenić zdolność kredytową danej osoby i podjąć decyzję o udzieleniu kredytu, np. karty kredytowej, **kliknij poniższy obrazek**. Zostanie wyświetlona symulacja oceny zdolności kredytowej, która pozwoli ocenić prawdopodobieństwo uzyskania karty kredytowej.

Wyzwania i zagrożenia związane z stronicznych i niereprezentatywnych danych

Obawy dotyczące wyjaśnialności

Chociaż modele uczenia maszynowego służące do oceny zdolności kredytowej i udzielania kredytów mają wiele zalet, istnieją również poważne obawy dotyczące przejrzystości, sprawiedliwości i potencjalnego wzmocnienia stronniczości, co z kolei może prowadzić do wykluczenia kredytowego, a nie włączenia.



Jedną z konkretnych obaw jest **brak wyjaśnialności** lub niemożność wyjaśnienia, dlaczego programy uczenia maszynowego podejmują określone decyzje po analizie wprowadzonych danych. Stanowi to problem dla użytkowników systemów, organów regulacyjnych i stron trzecich, które mogą nie być w stanie zrozumieć i wyjaśnić, dlaczego program podjął daną decyzję i dlaczego. Zdolność sztucznej inteligencji do „reagowania na duże ilości różnorodnych danych wejściowych, przekraczających możliwości poznawcze człowieka, jest również **piętą achillesową uczenia maszynowego**, ponieważ **taka złożoność jest często nieprzejrzysta pod względem procesu decyzyjnego poprzedzającego podjęcie decyzji**”.⁸ Uczenie maszynowe, które podejmuje tak złożone decyzje, że nie są one łatwe do zinterpretowania lub wyjaśnienia przez ludzi, jest zwykle określane jako „czarna skrzynka ML” lub „czarna skrzynka AI”.⁹

Stanowi to poważne wyzwanie, ponieważ **ogranicza przejrzystość i odpowiedzialność**. Gdy wnioskodawcy odmawia się udzielenia kredytu, kredytodawca powinien być w stanie wyjaśnić powód odmowy. Jeśli kredytodawcy nie są w stanie wystarczająco wyjaśnić, w jaki sposób zautomatyzowane systemy podjęły decyzję, wnioskodawcy mogą nie mieć wystarczających argumentów, aby zakwestionować tę decyzję i mogą czuć się bezbronni, co podważa zaufanie do

samej instytucji finansowej. Ponadto kredytodawcy muszą mieć pewność, że są w stanie uzasadnić swoje decyzje, aby zachować zgodność z przepisami, które mogą wymagać, aby decyzje dotyczące kredytów i pożyczek miały jasne motywy, które należy ujawnić konsumentom.¹⁰ Kolejnym problemem jest to, że **Black Box AI** utrudnia działania mające na celu ulepszenie systemu. Jeśli decyzja jest nieodpowiednia, „**niezwykle trudno jest przeanalizować, dlaczego popełniono błąd, lub określić, co należy zrobić, aby skorygować model**”.¹¹

Możliwość wyjaśnienia decyzji podejmowanych przez sztuczną inteligencję ma również zasadnicze znaczenie dla zdobycia zaufania użytkowników i zapewnienia, że decyzje te są sprawiedliwe i uczciwe, dlatego też **niezbędne stają się metody kontroli tych systemów**. Wyjaśnialność powinna zazwyczaj koncentrować się na wyjaśnieniu, dlaczego „ten konkretny wkład doprowadził do tego konkretnego wyniku”¹², ale istotna jest również wiedza na temat tego, jakie dane wewnętrzne tworzą struktury danego programu. W poniższej sekcji skupiamy się na kwestiach związanych z danymi wykorzystywanymi do szkolenia systemów, które ostatecznie są łączone z konkretnymi danymi przekazanymi przez użytkownika i umożliwiają systemowi AI podjęcie konkretnej decyzji.

8 King's College London. (b.d.). Wyzwania związane z wyjaśnialnością sztucznej inteligencji.

<https://www.kcl.ac.uk/challenges-of-ai-explainability>

9 Kosinski, M. (29 października 2024 r.). Czym jest sztuczna inteligencja typu „czarna skrzynka” i jak działa? IBM.

<https://www.ibm.com/think/topics/black-box-ai>

10 Congressional Research Service. (2023). Generatywna sztuczna inteligencja: przegląd, problemy i pytania dla Kongresu (raport CRS nr IF12399). <https://crsreports.congress.gov/product/pdf/IF/IF12399>

11 King's College London. (b.d.). Wyzwania związane z wyjaśnialnością sztucznej inteligencji.

<https://www.kcl.ac.uk/challenges-of-ai-explainability>

12 Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M. i Kagal, L. (2018). Wyjaśnianie wyjaśnień: przegląd interpretowalności uczenia maszynowego. ArXiv. <https://arxiv.org/abs/1806.00069>

Brak różnorodności zbiorów danych i wadliwe dane

Jak wyjaśniono wcześniej, stronnicze dane i algorytmy mogą sprawić, że automatyczne podejmowanie decyzji prowadzi do wyników, które stawiają w niekorzystnej sytuacji osoby mające trudności z uzyskaniem dobrego kredytu, a nawet dostępu do usług finansowych. Mniejszości i grupy o niskich dochodach zazwyczaj cierpią z powodu tych stronniczości w nieproporcjonalny sposób.

Jednak badania pokazują, że nie jest to jedyny problem. Różne wyniki dla mniejszości i większości nie są związane wyłącznie z tendencyjnością, ale także z faktem, że mniejszości i grupy o niskich dochodach mają mniej danych w swojej historii kredytowej, ponieważ zazwyczaj mają ograniczony dostęp do kredytów¹³. Oznacza to, że gdy dane te są wykorzystywane do obliczenia oceny kredytowej, a ocena ta jest wykorzystywana do przewidywania niewypłacalności kredytobiorcy, przewidywania te będą mniej precyzyjne. To właśnie ten brak precyzji prowadzi do nierówności, a nie tylko stronniczość¹⁴. Brak różnorodności w zbiorach danych służących do szkolenia modeli uczenia maszynowego powoduje szkody dla określonych społeczności, co zwiększa nierówności. Systemowe nierówności utrzymują się w zakresie selekcji zbiorów danych i nierówności dostępu. W

niektórych dziedzinach wynika to również z nierównych możliwości uczestniczenia w tworzeniu tych zbiorów danych¹⁵. Wydajność każdego systemu sztucznej inteligencji jest w dużym stopniu uzależniona od zbiorów danych, które analizuje za pomocą statystyki, ponieważ jego wyniki pochodzą z identyfikacji wzorców w danych: „Jakość zbioru danych leżącego u podstaw systemu sztucznej inteligencji ma kluczowe znaczenie dla jego skuteczności”.¹⁶ W kontekście udzielania kredytów i dostępu do kredytów „jest to cykl samonapędzający się... Udzielamy kredytów niewłaściwym osobom, a znaczna część społeczeństwa nigdy nie ma szansy zgromadzić danych potrzebnych do uzyskania kredytu w przyszłości”.¹⁷

Ryzyko stronniczości algorytmicznej

Chociaż jedną z zalet automatycznej analizy wniosków kredytowych może być zmniejszenie subiektywności w procesie podejmowania decyzji o przyznaniu kredytu, istnieje ryzyko, że procesy te utrwalają „istniejące uprzedzenia i stereotypy wobec grup definiowanych przez rasę, płeć, orientację seksualną i inne cechy”¹⁸, z których niektóre stanowią szczególne kategorie danych osobowych chronionych przez prawo.¹⁹

Wynika to z faktu, że zbiory danych zazwyczaj zawierają decyzje podjęte w przeszłości przez instytucje finansowe lub nie zawierają wystarczających danych dotyczących niektórych grup, co może prowadzić do ich dyskryminacji. Klikając na poniższy obrazek, można uzyskać dostęp do artykułu wyjaśniającego, w jaki sposób automatyczny system zatwierdzania kart

kredytowych Apple Card, dostarczany przez Goldman Sachs, doprowadził do skarg dotyczących domniemanej dyskryminacji kobiet ubiegających się o kartę, które twierdziły, że oferowano im niższe limity kredytowe lub odmawiano wydania karty, nawet jeśli ich mężowie otrzymali zatwierdzenie i lepsze warunki.



- 13 Blattner, L. i Nelson, S. (2021). Jak kosztowne jest szumy? Dane i dysproporcje w kredytach konsumenckich. ArXiv. <https://arxiv.org/abs/2105.07554>
- 14 Heaven, W. D. (2021). Stronniczość nie jest jedynym problemem związanym z ocenami kredytowymi – i nie, sztuczna inteligencja nie może w tym pomóc. MIT Technology Review. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>
- 15 Arora, A., Alderman, J. E., Palmer i in. (2023). Wartość standardów dla zbiorów danych dotyczących zdrowia w aplikacjach opartych na sztucznej inteligencji. BMJ Health & Care Informatics, 30(1), e100888. <https://doi.org/10.1136/bmjhci-2023-100888>
- 16 Bach, A. i Kriebitz, A. (29 września 2023 r.). Różnorodność w sztucznej inteligencji – w kierunku sformułowania problemu. Human Technology Foundation. <https://www.human-technology-foundation.org/news/diversity-in-ai-towards-a-problem-statement>
- 17 Heaven, W. D. (2021). Stronniczość nie jest jedynym problemem związanym z ocenami kredytowymi – i nie, sztuczna inteligencja nie może w tym pomóc. MIT Technology Review. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>
- 18 Garcia, A. C. B., Garcia, M. G. P. i Rigobon, R. (2024). Dyskryminacja algorytmiczna w dziedzinie kredytów: co o niej wiemy? AI & Society, 39, 2059–2098. <https://doi.org/10.1007/s00146-023-01676-3>
- 19 Artykuł 9 ogólnego rozporządzenia o ochronie danych: <https://gdpr-info.eu/art-9-gdpr/>

Odmowa dostępu do rynków kredytowych może być dyskryminująca, gdy decyzje są wypaczone na podstawie istniejących uprzedzeń. Może to również przejawiać się w różnicowym traktowaniu, które oferuje gorsze warunki osobom cierpiącym z powodu dyskryminacji, takie jak inne, mniej korzystne opłaty lub wyższe stopy procentowe.²⁰

Uprzedzenia w zbiorach danych mogą **stworzyć dodatkowe bariery w dostępie do usług finansowych dla tradycyjnie niedostatecznie obsługiwanych grup społecznych**, które już wcześniej miały trudności z dostępem do tych usług. Tradycyjne systemy oceny zdolności kredytowej penalizują osoby bez formalnej historii kredytowej, ale nawet systemy wykorzystujące nietradycyjne dane, jeśli wykazują uprzedzenia w swoich decyzjach lub wzmacniają uprzedzenia swoich twórców, mogą **pogłębiać wykluczenie**, wywierając nieproporcjonalny wpływ na te

niedostatecznie obsługiwane społeczności.

Ograniczony dostęp niektórych grup społecznych do tradycyjnych usług finansowych może utrwaląć dysproporcje w zamożności. Dyskryminacja w tej dziedzinie w przeszłości miała długoterminowy wpływ na ogólną zamożność niektórych grup, powodując takie konsekwencje, jak „ograniczona możliwość nabywania nieruchomości, zmniejszone możliwości przedsiębiorcze i pokoleniowe różnice w zamożności”.²¹

Jeśli **modele sztucznej inteligencji** nie zostaną starannie zaprojektowane i wdrożone, **będą „utrwaląć nierówności”**²², niwelując wszelkie możliwe pozytywne skutki procesu decyzyjnego, który ogranicza subiektywność tylko na papierze, ponieważ niesie ze sobą strukturalne uprzedzenia już obecne w społeczeństwie.

Skąd biorą się błędy systemowe?

Błędy algorytmiczne mogą wynikać z różnych źródeł, **w tym z tendencyjnych danych szkoleniowych**, które mogą być danymi historycznymi odzwierciedlającymi istniejące uprzedzenia społeczne. Jest to znane jako „bias in, bias out” lub sytuacja, w której model jest szkolony na danych związanych z już tendencyjnymi wynikami. Sztuczna inteligencja może po prostu powielać problemy z przeszłości.

Brak różnorodności jest widoczny nie tylko w danych. Może on również występować w **zespołach projektujących produkty**: „Jednorodność wśród analityków danych i programistów przyczynia się do utrwalania stronniczości w systemach sztucznej inteligencji”.²³ Brak doświadczenia lub zrozumienia dla społeczności znajdujących się w niekorzystnej sytuacji może sprawić, że osoby projektujące systemy i dostarczające dane szkoleniowe będą ślepe na ich trudności i wyzwania lub specyficzne sytuacje, co może prowadzić do niedokładnej analizy ich historii kredytowej.

Stronnicze decyzje **mogą z czasem tylko się utrwalać**, ponieważ systemy dokonują coraz więcej błędnych ocen, które są następnie uznawane za prawidłowe, generując **pętle sprzężenia zwrotnego**, w których błędne odmowy udzielenia kredytu określonej grupie znajdują odzwierciedlenie w **przyszłych danych szkoleniowych, co jeszcze bardziej utrwalą nierówności i niesprawiedliwość**.²⁴

20 Garcia, A. C. B., Garcia, M. G. P. i Rigobon, R. (2024). Dyskryminacja algorytmiczna w dziedzinie kredytów: co o niej wiemy? *AI & Society*, 39, 2059–2098. <https://doi.org/10.1007/s00146-023-01676-3>

21 Nuka, T. F. i Osedahunsi, B. O. (2024). Od stronniczości do równowagi: integracja DEI w systemach finansowych opartych na sztucznej inteligencji w celu promowania równości kredytowej. *International Journal of Science and Research Archive*, 13(2), 1189–1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

22 Nuka, T. F. i Osedahunsi, B. O. (2024). Od stronniczości do równowagi: integracja DEI w systemach finansowych opartych na sztucznej inteligencji w celu promowania równości kredytowej. *International Journal of Science and Research Archive*, 13(2), 1189–1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

23 Nuka, T. F. i Osedahunsi, B. O. (2024). Od stronniczości do równowagi: integracja DEI w systemach finansowych opartych na sztucznej inteligencji w celu promowania równości kredytowej. *International Journal of Science and Research Archive*, 13(2), 1189–1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

24 Nuka, T. F. i Osedahunsi, B. O. (2024). Od stronniczości do równowagi: integracja DEI w systemach finansowych opartych na sztucznej inteligencji w celu promowania równości kredytowej. *International Journal of Science and Research Archive*, 13(2), 1189–1206. <https://doi.org/10.30574/ijrsra.2024.13.2.2257>

W sprawie studium przypadku

Niniejsze studium przypadku daje wykładowcom możliwość zainicjowania dyskusji na temat kilku zagadnień związanych z integracją i wykluczeniem finansowym, w tym potencjału i ryzyka związanego z automatyczną oceną zdolności kredytowej i automatycznymi decyzjami kredytowymi.

Najpierw można rozpocząć od dyskusji na temat oceny zdolności kredytowej opartej na sztucznej inteligencji oraz usług finansowych opartych na sztucznej inteligencji, omawiając zalety automatyzacji i potencjał wykorzystania różnych rodzajów zbiorów danych do obsługi różnych grup społecznych oraz skupiając się na integracji.

Po zapoznaniu się z procesem zatwierdzania kredytów należy skupić się na wykorzystaniu danych osobowych do podejmowania tego rodzaju decyzji, ponieważ modele oparte na sztucznej inteligencji wymagają zbiorów danych nie tylko do szkolenia, ale także do wprowadzania danych osobowych przez użytkowników, jak pokazano w symulacji na stronie 8 niniejszego studium przypadku.

Oto kilka pytań dla uczniów:



- Czy uważasz, że istnieje ryzyko dla prywatności ludzi i ochrony danych osobowych?
- Jeśli tak, dlaczego uważasz, że ochrona danych osobowych jest ważna?
- Jakie konsekwencje mogą ponieść klienci, jeśli ich dane zostaną ujawnione?
- Oprócz wycieków danych i innych zagrożeń dla cyberbezpieczeństwa, jakie inne zagrożenia możecie powiązać z wykorzystaniem danych osobowych do podejmowania decyzji kredytowych?

Po przejściu do części czwartej należy omówić z uczniami poszczególne wyzwania, a następnie skupić się na rozwiązaniach. Studium przypadku zawiera kilka odniesień i dalszych lektur w załącznikach poniżej, ale ciekawym ćwiczeniem jest przydzielenie uczniów do grup i zlecenie im przeprowadzenia badań w celu opracowania możliwych rozwiązań dla każdego z przedstawionych problemów. Jedna lub więcej grup może skupić się na badaniach dotyczących kwestii wyjaśnialności, inne na rozwiązaniach związanych z problemem różnorodności zbiorów danych, a jeszcze inne na ogólnych tendencyjnościach algorytmicznych. Po przeprowadzeniu badań można poprosić uczniów o przygotowanie krótkiej prezentacji, w której przedstawiają rozwiązania, które mogliby wdrożyć.

Można polecić uczniom poszukiwanie rozwiązań, od metod regulacyjnych i legislacyjnych po rozwiązania bardziej techniczne. Istnieją świetne zasoby internetowe poświęcone tym tematom, więc jest to również dobra okazja, aby uczniowie przeprowadzili

badania, podczas których będą krytycznie myśleć i identyfikować wiarygodne źródła proponujące wartościowe rozwiązania.

Na koniec, jeśli uznasz, że pasuje to do kontekstu Twoich zajęć, możesz omówić zalety i pułapki ogólnej tendencji do automatyzacji procesu podejmowania decyzji, która może mieć wpływ na życie ludzi, a także potrzebę, aby przynajmniej ludzie nadzorowali te decyzje. Możesz zainicjować interesującą dyskusję na temat potrzeby interakcji międzyludzkiej i przyznawania wnioskodawcom możliwości uzyskania kredytu w porównaniu z zaletami decyzji wspomaganych przez automatyzację lub nawet delegowanych do systemów sztucznej inteligencji oraz potencjałem decyzji eliminujących subiektywność. Czy jest to pożądane? Czy jest to w ogóle możliwe? A może wręcz przeciwnie, pewien stopień subiektywności i ludzkiej sprawczości jest niezbędny do podejmowania sprawiedliwych decyzji?



Więcej informacji

- Alliance for Financial Inclusion. (2025). Alternative data for credit scoring. <https://www.afi-global.org/wp-content/uploads/2025/02/Alternative-Data-for-Credit-Scoring.pdf>
- Alonso, A., & Carbó, J. M. (2022). Accuracy of explanations of machine learning models for credit decisions (Working Paper No. 2222). Banco de España. <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSerias/DocumentosTrabajo/22/Files/dt2222e.pdf>
- Blattner, L., & Nelson, S. (2022). How flawed data aggravates inequality in credit. Stanford Graduate School of Business. <https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>
- Dash, R., Kremer, A., & Petrov, A. (2021). Designing next-generation credit-decisioning models. McKinsey & Company. <https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/designing-next-generation-credit-decisioning-models>
- Heaven, W. D. (2021). Bias isn't the only problem with credit scores – and no, AI can't help. MIT Technology Review. <https://www.technologyreview.com/2021/06/17/1026519/racial-bias-noisy-data-credit-scores-mortgage-loans-fairness-machine-learning/>
- Kelion, L. (2019). Apple's 'sexist' credit card investigated by US regulator. BBC News. <https://www.bbc.com/news/business-50365609>
- Mensalvas, E., Guzmán, M. A. & Ruiz Bonilla, S. (2022). ML applied to credit risk: Building explainable models. <https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2022/10/Idanae-3Q22.pdf>
- ODSC. (2022). Should AI decide who gets a loan? Medium. <https://odsc.medium.com/should-ai-decide-who-gets-a-loan-83c6f259081b>