



www.aileaders-project.eu

SYMULACJA – Wyzwania etyczne w zakresie zasobów ludzkich



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.

SYMULACJA

- Wyzwania etyczne w zakresie zasobów ludzkich

- 01  Streszczenie 3
- 02  Wprowadzenie 4
- 03  Prezentacja narzędzi 6
- 04  Wykonanie symulacji 7
- 05  Wnioski 8
- 06  Bibliografia 10

• 01 Streszczenie



Rodzaj OER

Demonstracja/symulacja przy użyciu Google Colab (RandomForest vs FairGBM Classifier)

Cel/przeznaczenie

Zapewnienie studentom praktycznej i krytycznej analizy tego, w jaki sposób algorytmiczne podejmowanie decyzji w procesie rekrutacji może reprodukować nierówności strukturalne poprzez porównanie wyników uzyskanych na podstawie wskaźników sprawiedliwości. Symulacja ta zachęca do refleksji nad etycznymi aspektami sztucznej inteligencji w zasobach ludzkich i promuje opracowywanie bardziej sprawiedliwych i integracyjnych modeli predykcyjnych.

Oczekiwane efekty kształcenia

Pod koniec symulacji studenci będą potrafili:

- 01** Wykrywać i interpretować algorytmiczne uprzedzenia w rekrutacji opartej na sztucznej inteligencji;
- 02** Poddać się refleksji nad etycznymi implikacjami zautomatyzowanych systemów rekrutacyjnych
- 03** Wykorzystać wskaźniki sprawiedliwości do oceny wyników modelu;

Słowa kluczowe

- Uczenie maszynowe
- Klasyfikacja
- Zasoby ludzkie
- Uprzedzenia
- Sprawiedliwość

Sugerowane podejście metodologiczne

Nauka oparta na problemach

• 02 Wprowadzenie



Słownik

- 01** **Odpowiedzialność** oznacza, że organizacja ponosi odpowiedzialność za swoje decyzje dotyczące zatrudnienia i ich skutki, zapewniając konsekwencje za nieuczciwe praktyki.
- 02** **Przejrzystość** polega na tym, że proces rekrutacji i kryteria wyboru są jasne i zrozumiałe dla wszystkich kandydatów.
- 03** **Sprawiedliwość** gwarantuje, że wszyscy kandydaci są oceniani bezstronnie na podstawie ich kwalifikacji do danego stanowiska, bez żadnych uprzedzeń lub dyskryminacji.
- 04** **Uprzedzenia społeczne** w zatrudnianiu to nieuczciwa tendencja do faworyzowania lub odrzucania kandydatów na podstawie uprzedzeń dotyczących ich grupy społecznej (np. wieku, płci lub rasy) zamiast ich rzeczywistych umiejętności i kwalifikacji.

Coraz częściej oczekuje się, że procesy rekrutacyjne będą wspierane przez technologie sztucznej inteligencji (AI), szczególnie na wczesnych etapach zatrudniania, takich jak przeglądanie CV, ranking kandydatów i tworzenie listy osób zakwalifikowanych do rozmów kwalifikacyjnych. Chociaż systemy te oferują potencjał większej wydajności i skalowalności, budzą one również ważne obawy etyczne, szczególnie dotyczące sprawiedliwości i stronniczości (Horodyski, 2023).

Na długo przed wprowadzeniem sztucznej inteligencji badania nad tradycyjnymi praktykami rekrutacyjnymi i selekcyjnymi wykazały już utrzymujące się wzorce dyskryminacji (Breaugh, 2013; Hebl et al., 2020). Obejmują one uprzedzenia związane z rasą, pochodzeniem etnicznym i statusem mniejszości (np. Allen & Vardaman, 2017; Hiemstra et al., 2013; Veit & Thijsen, 2019; Zschirnt & Ruedin, 2016); płeć (np. Ellemers, 2018); klasa społeczna (np. Henderson, 2018); wiek (np. Czopp et al., 2015; Zaniboni i in., 2019); oraz wykształcenie (np. Daly i in., 2000).

Coraz częściej oczekuje się, że procesy rekrutacyjne będą wspierane przez technologie sztucznej inteligencji (AI), szczególnie na wczesnych etapach zatrudniania, takich jak przeglądanie CV, ranking kandydatów i tworzenie listy osób zakwalifikowanych do rozmów kwalifikacyjnych, co może budzić istotne obawy etyczne, zwłaszcza dotyczące sprawiedliwości i stronniczości (Horodyski, 2023).

Na długo przed wprowadzeniem sztucznej inteligencji badania nad tradycyjnymi praktykami rekrutacji i selekcji wykazały już utrzymujące się wzorce dyskryminacji (Hebl i in., 2020). Obejmują one uprzedzenia związane z rasą, wiekiem, pochodzeniem etnicznym i statusem mniejszości (np. Allen i Vardaman, 2017; Veit i Thijsen, 2019; Zaniboni i in., 2019).



Obecnie istnieje obawa, że technologie sztucznej inteligencji, zamiast łagodzić te uprzedzenia, mogą je utrzymywać, a nawet wzmacniać. Najnowsze badania sugerują, że algorytmiczne podejmowanie decyzji w zakresie zasobów ludzkich może powielać historyczne nierówności zakodowane w danych wykorzystywanych do szkolenia takich systemów (np. Rigotti & Fosh-Villaronga, 2024; Seppälä & Małecka, 2024). W rezultacie wykorzystanie sztucznej inteligencji w rekrutacji rodzi krytyczne pytania dotyczące odpowiedzialności, przejrzystości i sprawiedliwości w praktykach zatrudniania.

W tym kontekście niniejsze ćwiczenie symulacyjne ma na celu zaangażowanie studentów w etyczne wyzwania związane z rekrutacją i selekcją kandydatów opartą na sztucznej inteligencji. Celem jest zapewnienie praktycznego doświadczenia w identyfikowaniu potencjalnych uprzedzeń w algorytmach rekrutacyjnych oraz

badaniu strategii uwzględniających sprawiedliwość, aby wspierać bardziej integracyjne i sprawiedliwe podejmowanie decyzji.

Korzystając ze wskaźników sprawiedliwości, studenci przeanalizują, czy modele uczenia maszynowego wykorzystywane do przewidywania decyzji dotyczących wstępnej selekcji kandydatów odzwierciedlają – lub odtwarzają – historyczne wzorce dyskryminacji, ze szczególnym uwzględnieniem pochodzenia etnicznego, płci i statusu społeczno-ekonomicznego.

Stawiając czoła tym wyzwaniom w kontrolowanym, symulowanym środowisku, studenci rozwiną zarówno kompetencje techniczne, jak i wrażliwość etyczną niezbędną do odpowiedzialnego projektowania i wdrażania sztucznej inteligencji w zarządzaniu zasobami ludzkimi.



• 03 Prezentacja narzędzi



Ta symulacja dotyczy zadania klasyfikacji binarnej w rekrutacji opartej na sztucznej inteligencji: przewidywania, czy kandydat powinien zostać zakwalifikowany do dalszego etapu rekrutacji na stanowisko kierownicze średniego szczebla na podstawie swojego profilu.

Wykorzystany zbiór danych jest syntetyczny, ale wzorowany na rzeczywistych danych rekrutacyjnych, publicznie dostępnych za pośrednictwem platformy Kaggle. Atrybuty kandydatów obejmują płeć, rasę/pochodzenie etniczne, poziom wykształcenia, lata doświadczenia, oczekiwania płacowe, źródło rekrutacji.

Chociaż zbiór danych jest sztuczny, realistycznie symuluje typowe decyzje i wzorce rekrutacyjne, dzięki czemu idealnie nadaje się do badania sprawiedliwości w uczeniu maszynowym. Szczególną uwagę zwraca się na sposób, w jaki wrażliwe zmienne, takie jak rasa/pochodzenie etniczne, mogą wpływać na prognozy modelu.

Badania wykazały, że procesy rekrutacyjne mogą systematycznie stawiać w niekorzystnej sytuacji kandydatów z określonych środowisk etnicznych, nawet jeśli mają oni takie same kwalifikacje (Zschirnt & Ruedin, 2016; Veit & Thijsen, 2019; Hiemstra et al., 2013).

Ta symulacja koncentruje się na jednym **głównym pytaniu**:

Czy algorytmy uczenia maszynowego w nieproporcjonalny sposób „zalecają” odrzucenie podań o pracę złożonych przez osoby czarnoskóre i kobiety?

Badając tę kwestię, studenci krytycznie oceniają, czy modele sztucznej inteligencji stosowane w selekcji kandydatów wykazują niezamierzone uprzedzenia oraz w jaki sposób wzorce te odnoszą się do empirycznych wyników badań przedstawionych w literaturze naukowej.

04 Wykonanie symulacji



01 Dostęp do zeszytu symulacji

Przejdź do strony <https://tinyurl.com/k63793wp>

02 Uruchom kod

- Aby wykonać wszystkie komórki, kliknij kartę „Runtime” i wybierz opcję „Run All”. Poczekać chwilę (1 min).
- Upewnij się, że wyniki ładują się poprawnie i że wszystkie modele zostały pomyślnie wytrenowane.

03 Przeglądaj zbiór danych

- W narzędziu What If Tool kliknij „Performance and Metrics” (Wydajność i wskaźniki), a następnie wybierz funkcję „Over 50k” (Ponad 50 tys.) w sekcji „Ground Truth Feature” (Funkcja Ground Truth) i wybierz „GenderCode” (Kod płci) dla opcji „Slice by” (Podziel według).
- Przejrzyj zbiór danych i sprawdź kluczowe zmienne. Zwróć szczególną uwagę na cechy demograficzne, takie jak rasa/pochodzenie etniczne i płeć, oraz ich potencjalną korelację z wynikami selekcji.

04 Analiza sprawiedliwości

Zastosuj wskaźniki sprawiedliwości, aby przeanalizować różnice we wskaźnikach dokładności odrzucenia w różnych grupach demograficznych, a mianowicie rasie/pochodzeniu etnicznym.

Datapoint editor

Performance & Fairness

Features

600 datapoints loaded

?

Configure

Ground Truth Feature

Over-50K

WHAT IS GROUND TRUTH?

The feature that your model is trying to predict. [More](#).

Cost Ratio (FP/FN)

1

WHAT IS COST RATIO?

The cost of false positives relative to false negatives. Required for optimization. [More](#).

Slice by

RaceDesc

WHAT DOES SLICING DO?

Shows the model's performance on datapoints grouped by each value of the selected feature.

Slice by (secondary)

<none>

Fairness

Apply an optimization strategy

Select a strategy to automatically set classification thresholds, based on the set cost ratio and data slices. Manually altering thresholds or changing cost ratio will revert the strategy to 'custom thresholds'.

☒ Custom thresholds

☐ Single threshold

☐ Demographic parity

☐ Equal opportunity

☐ Equal accuracy

☐ Group thresholds

Custom thresholds for 5 values of RaceDesc

Sort by

Count

Count

False Positives (%)

False Negatives (%)

Accuracy (%)

F1

Feature Value	Count	Model	Threshold	False Positives (%)	False Negatives (%)	Accuracy (%)	F1
▶ Hispanic	125	1		0.5	0.0	24.0	76.0
	2			0.5	3.2	20.8	76.0
▶ Other	124	1		0.5	0.0	28.2	71.8
	2			0.5	3.2	25.0	71.8
▶ Black	122	1		0.5	0.0	23.8	76.2
	2			0.5	4.9	18.9	76.2
▶ Asian	117	1		0.5	0.0	24.8	75.2
	2			0.5	5.1	19.7	75.2
▶ White	112	1		0.5	0.9	28.6	70.5
	2			0.5	8.0	21.4	70.5

05 Porównaj wyniki

Porównaj wyniki w różnych grupach, zwłaszcza pod względem pochodzenia etnicznego. Sprawdź, czy kandydaci o takich samych kwalifikacjach, ale z różnych środowisk etnicznych, są traktowani inaczej. Zastanów się, jak to się ma do znanych schematów dyskryminacji przy zatrudnianiu.

06 Zastanów się nad implikacjami etycznymi

Omów, czy i w jaki sposób model odzwierciedla lub wzmacnia społeczne uprzedzenia. Oceń skuteczność interwencji uwzględniających kwestię sprawiedliwości. Rozważ szersze implikacje etyczne związane z wykorzystaniem sztucznej inteligencji w rekrutacji i zarządzaniu talentami.

• 05 Wnioski



Ta symulacja ilustruje, w jaki sposób modele uczenia maszynowego, jeśli nie są kontrolowane, mogą utrwalać i reprodukować historyczne uprzedzenia w zautomatyzowanych procesach rekrutacyjnych. Chociaż wykorzystany zbiór danych jest syntetyczny, a model celowo uproszczony, wyniki odzwierciedlają rzeczywiste obawy dotyczące sprawiedliwości w zatrudnianiu, szczególnie w odniesieniu do różnic płciowych i etnicznych.

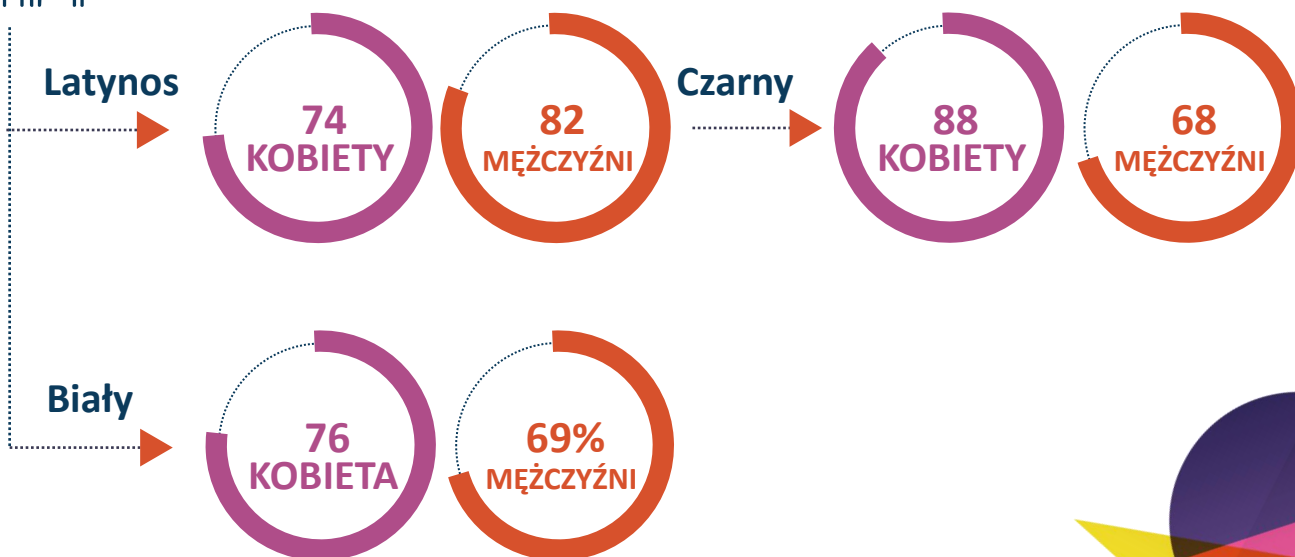
Wyniki ujawniają znaczne niespójności w dokładności prognoz w różnych grupach demograficznych. W szczególności dokładność prognoz dotyczących odrzucenia kandydatury jest znacznie wyższa w przypadku czarnych kobiet. W

tym kontekście wyższa dokładność odpowiada niższemu wskaźnikowi fałszywych wyników negatywnych, co oznacza, że aplikacje czarnych kobiet są częściej i prawidłowo klasyfikowane jako odrzucone.

Podsumowanie wyników



Dokładność w przewidywaniu decyzji o odrzuceniu kandydatów — tj. prawdziwe „wyniki negatywne” (%)



Sugeruje to, że model w nieproporcjonalny sposób „zaleca” odrzucanie kandydatów z tej grupy w porównaniu z kandydatami z innych grup, co wskazuje na niepokojącą stronniczość w procesie podejmowania decyzji przez algorytm.

Odkrycia te ponownie wywołują krytyczną dyskusję na temat słabych punktów algorytmów rekrutacyjnych, w szczególności sposobów, w jakie nierówności strukturalne są kodowane w danych szkoleniowych, a następnie przyswajane przez systemy sztucznej inteligencji. Obserwacja, że dokładność prognozowania różni się w zależności od grup marginalizowanych — a nawet w obrębie kategorii chronionych — podkreśla złożoność oceny sprawiedliwości: większa dokładność niekoniecznie oznacza bardziej sprawiedliwe wyniki.

Stosując wskaźniki sprawiedliwości i analizując wyniki na poziomie grupowym, studenci są zachęceni do krytycznego podejścia do społeczno-technicznych wymiarów stronniczości algorytmicznej. Ćwiczenie to nie tylko zapewnia praktyczne doświadczenie w zakresie wykrywania stronniczości i technik jej ograniczania, ale także podkreśla etyczną odpowiedzialność programistów i analityków za projektowanie sprawiedliwych i odpowiedzialnych systemów sztucznej inteligencji. W kontekście zarządzania zasobami ludzkimi sprawiedliwość nie jest jedynie kwestią prawną lub reputacyjną — jest to podstawowy obowiązek etyczny.

• 06 Bibliografia



- Allen, D. G., & Vardaman, J. M. (2017). Recruitment and Retention Across Cultures. *Annual Review of Organizational Psychology and Organizational Behavior*, 4(1), 153–181.
- Breugh, J. A. (2013). Employee Recruitment. *Annual Review of Psychology*, 64(1), 389–416.
- Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive Stereotypes Are Pervasive and Powerful. *Perspectives on Psychological Science*, 10(4), 451–463.
- Daly, M. C., Büchel, F., & Duncan, G. J. (2000). Premiums and penalties for surplus and deficit education: Evidence from the United States and Germany. *Economics of Education Review*, 19(2), 169–178.
- Ellemers, N. (2018). Gender Stereotypes. *Annual Review of Psychology*, 69(1), 275–298.
- HeModern Discrimination in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, M., Cheng, S. K., & Ng, L. C. (2020)., 7(1), 257–282.
- Henderson, D. (2018). The Effects of Social Class on Perceptions of Job Applicants' Suitability for Employment. *Academy of Management Proceedings*, 2018(1), 13748.
- Hiemstra, A. M. F., Derous, E., Serlie, A. W., & Born, M. P. (2013). Ethnicity Effects in Graduates' Résumé Content. *Applied Psychology*, 62(3), 427–453.
- Horodyski, P. (2023). Recruiter's perception of artificial intelligence (AI)-based tools in recruitment. *Computers in Human Behavior Reports*, 10, 100298.
- Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review*, 53, 105966.
- Seppälä, P., & Małecka, M. (2024). AI and discriminative decisions in recruitment: Challenging the core assumptions. *Big Data & Society*, 11(1), 20539517241235872
- Veit, S., & Thijsen, L. (2019). Almost identical but still treated differently: hiring discrimination against foreign-born and domestic-born minorities. *Journal of Ethnic and Migration Studies*, 1–20.
- Zaniboni, S., Kmicinska, M., Truxillo, D. M., Kahn, K., Paladino, M. P., & Fraccaroli, F. (2019). Will you still hire me when I am over 50? The effects of implicit and explicit age stereotyping on resume evaluations. *European Journal of Work and Organizational Psychology*, 28(4), 453–467.
- Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7), 1115–1134.



aileaders

Śledź naszą podróż



www.aileaders-project.eu



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.