UNIWERSYTET WARSZAWSKI

WYDZIAŁ EKONOMII

KIERUNEK: INFORMATYKA / DATA SCIENCE

PRACA DYPLOMOWA

**ZASTOSOWANIE ALGORYTMÓW UCZENIA MASZYNOWEGO**

**W PREDYKCJI ROTACJI PRACOWNIKÓW**

Autor: Paweł Książyk

Nr albumu: [Numer]

Promotor: Maciej

[Tytuł naukowy, Katedra]

Warszawa, 2025

SPIS TREŚCI

STRESZCZENIE ................................................................. 5

ABSTRACT .................................................................... 6

1. WPROWADZENIE ............................................................ 7

1.1 Uzasadnienie wyboru tematu .......................................... 7

1.2 Problem badawczy i pytania badawcze ................................ 9

1.3 Cele pracy i hipotezy badawcze ..................................... 11

1.4 Struktura pracy .................................................... 13

2. PRZEGLĄD LITERATURY ..................................................... 15

2.1 Teoretyczne podstawy rotacji pracowników .......................... 15

2.1.1 Teoria dwuczynnikowa Herzberga ............................... 16

2.1.2 Model Job Demands-Resources (JD-R) ........................... 17

2.1.3 Teoria wymiany społecznej .................................... 18

2.2 HR Analytics i People Analytics ................................... 19

2.2.1 Ewolucja analityki w zarządzaniu zasobami ludzkimi .......... 19

2.2.2 Zastosowania ML w talent management .......................... 21

2.2.3 Etyczne aspekty AI w HR ...................................... 22

2.3 Algorytmy uczenia maszynowego w klasyfikacji ...................... 24

2.3.1 Modele liniowe: Logistic Regression .......................... 24

2.3.2 Modele ensemble: Random Forest, Extra Trees, XGBoost ...................... 25

2.3.3 Modele nieparametryczne: SVM, KNN ............................ 27

2.4 Cost-sensitive learning i optymalizacja biznesowa ................. 28

2.5 Identyfikacja luki badawczej ...................................... 30

3. METODOLOGIA BADANIA ..................................................... 32

3.1 Założenia metodologiczne ........................................... 32

3.2 Opis i charakterystyka danych ...................................... 33

3.3 Proces przetwarzania danych ........................................ 35

3.3.1 Data cleaning i preprocessing ................................ 35

3.3.2 Feature engineering .......................................... 37

3.3.3 Selekcja zmiennych ........................................... 39

3.4 Wybór i konfiguracja algorytmów ML ................................. 40

3.5 Metodologia ewaluacji i optymalizacji .............................. 42

3.5.1 Metryki wydajności ........................................... 42

3.5.2 Cross-validation i hyperparameter tuning .................... 43

3.5.3 Cost-sensitive threshold optimization ........................ 44

3.6 Narzędzia i środowisko badawcze .................................... 46

4. WYNIKI BADANIA .......................................................... 48

4.1 Eksploracyjna analiza danych ....................................... 48

4.1.1 Charakterystyka rozkładów zmiennych .......................... 48

4.1.2 Analiza korelacji i zależności ............................... 52

4.1.3 Identyfikacja wzorców attrition .............................. 55

4.2 Feature engineering i preprocessing ................................ 58

4.2.1 Tworzenie nowych zmiennych ................................... 58

4.2.2 Obsługa zmiennych kategorycznych ............................. 61

4.2.3 Skalowanie i normalizacja .................................... 63

4.3 Wyniki modeli bazowych ............................................. 65

4.3.1 Porównanie 5 algorytmów ML ................................... 65

4.3.2 Ranking modeli według AUC-ROC ................................ 68

4.3.3 Analiza overfittingu ......................................... 70

4.4 Hyperparameter tuning .............................................. 72

4.4.1 Optymalizacja Logistic Regression ............................ 72

4.4.2 Porównanie baseline vs tuned models .......................... 74

4.4.3 Wybór najlepszego modelu ..................................... 76

4.5 Cost-sensitive optimization ........................................ 78

4.5.1 Definicja kosztów biznesowych ................................ 78

4.5.2 Analiza progów decyzyjnych ................................... 80

4.5.3 Optymalizacja ROI ............................................ 82

4.6 Analiza feature importance ......................................... 85

4.6.1 TOP 10 najważniejszych czynników ............................. 85

4.6.2 Kategoryzacja według teorii HR ............................... 87

4.6.3 Interpretacja biznesowa ...................................... 89

5. DYSKUSJA I INTERPRETACJA ................................................ 92

5.1 Weryfikacja hipotez badawczych ..................................... 92

5.2 Interpretacja w kontekście teorii HR ............................... 95

5.3 Porównanie z literaturą przedmiotu ................................. 98

5.4 Analiza stabilności i generalizacji ................................ 101

6. IMPLIKACJE PRAKTYCZNE ................................................... 104

6.1 Business case i analiza ROI ........................................ 104

6.2 Strategia implementacji ............................................. 107

6.3 Rekomendacje dla organizacji ....................................... 110

6.4 Zarządzanie ryzykiem i ograniczenia ................................ 113

7. OGRANICZENIA BADANIA .................................................... 116

7.1 Ograniczenia metodologiczne ........................................ 116

7.2 Ograniczenia techniczne ............................................ 118

7.3 Ograniczenia kontekstowe ........................................... 119

8. KIERUNKI PRZYSZŁYCH BADAŃ ............................................... 121

8.1 Rozszerzenia metodologiczne ........................................ 121

8.2 Nowe źródła danych ................................................. 123

8.3 Praktyczne rozszerzenia ............................................ 124

9. ZAKOŃCZENIE ............................................................. 126

9.1 Podsumowanie wyników ............................................... 126

9.2 Wkład do wiedzy .................................................... 128

9.3 Znaczenie dla przyszłości HR ...................................... 129

BIBLIOGRAFIA ............................................................... 131

ZAŁĄCZNIKI ................................................................. 137

Załącznik A: Szczegółowe wyniki modeli ................................. 137

Załącznik B: Kod źródłowy (kluczowe fragmenty) ......................... 140

Załącznik C: Dodatkowe wizualizacje .................................... 145

Załącznik D: Słownik zmiennych ......................................... 148

Załącznik E: Pipeline implementacyjny .................................. 150

SPIS TABEL ................................................................. 152

SPIS RYSUNKÓW .............................................................. 154

Spis treści

[**1. WPROWADZENIE** 10](#_Toc209709431)

[**1.1 Uzasadnienie wyboru tematu (2-3 strony)** 10](#_Toc209709432)

[**1.1.1 Znaczenie problemu rotacji pracowników w współczesnych organizacjach** 10](#_Toc209709433)

[**1.1.2 Koszty rotacji pracowników - analiza ekonomiczna** 10](#_Toc209709434)

[**1.1.3 Potencjał technologii i machine learning w HR** 12](#_Toc209709435)

[**1.1.4 Cost-sensitive learning jako innowacja w HR Analytics** 12](#_Toc209709436)

[**1.2 Problem badawczy i pytania badawcze (2-3 strony)** 13](#_Toc209709437)

[**1.2.1 Sformułowanie problemu głównego** 13](#_Toc209709438)

[**1.2.2 Pytania badawcze szczegółowe** 14](#_Toc209709439)

[**1.2.3 Kontekst teoretyczny i praktyczny** 15](#_Toc209709440)

[**1.3 Cele pracy i hipotezy badawcze (2-3 strony)** 15](#_Toc209709441)

[**1.3.1 Cel główny** 15](#_Toc209709442)

[**1.3.2 Cele szczegółowe** 16](#_Toc209709443)

[**1.3.3 Hipotezy badawcze** 17](#_Toc209709444)

[**1.3.4 Oczekiwany wkład do wiedzy** 18](#_Toc209709445)

S**TRESZCZENIE**

Problem rotacji pracowników stanowi jeden z kluczowych wyzwań współczesnych organizacji, generując znaczące koszty związane z rekrutacją, szkoleniem i utratą wiedzy organizacyjnej. Niniejsza praca przedstawia kompleksowe badanie zastosowania algorytmów uczenia maszynowego w predykcji odejść pracowników z innowacyjnym podejściem feature engineering oraz hyperparameter optimization.

Celem głównym badania było opracowanie modelu predykcyjnego zdolnego do skutecznego przewidywania rotacji pracowników przy jednoczesnej optymalizacji wydajności predykcyjnej. Wykorzystano dataset IBM HR Analytics zawierający 1,470 obserwacji pracowników z 35 zmiennymi opisującymi demografię, warunki pracy, satysfakcję i historię kariery, charakteryzujący się 16.1% wskaźnikiem rotacji.

Metodologia badania obejmowała dwuetapowy proces: podstawowy tuning z porównaniem czterech algorytmów uczenia maszynowego (Random Forest, XGBoost, Logistic Regression z regularyzacją L1, K-Nearest Neighbors) oraz zaawansowaną optymalizację hiperparametrów z wykorzystaniem GridSearchCV i 5-fold StratifiedKFold cross-validation. Przeprowadzono kompleksowy proces feature engineering, tworząc ponad 15 nowych zmiennych biznesowych, w tym interakcje (Age\_Satisfaction\_Interaction), wskaźniki kompozytowe (WorkLife\_Stress\_Score) oraz dewiacje względem grup referencyjnych.

Wyniki podstawowego tuningu wykazały, że najlepszym modelem okazał się XGBoost, osiągający najwyższe AUC-ROC na poziomie 0.8453 (±0.0222), jednak z lekkim przeuczeniem (gap: 0.0391). Logistic Regression z regularyzacją L1 osiągnęła porównywalny wynik AUC-ROC 0.8431 (±0.0195) przy jednoczesnej najlepszej generalizacji (gap: 0.0094) oraz najwyższym F1-score 0.5778. Analiza feature importance zidentyfikowała Age\_Satisfaction\_Interaction, WorkLife\_Stress\_Score oraz MonthlyIncome\_Deviation\_from\_Department jako najważniejsze predyktory rotacji.

Badanie potwierdza skuteczność dwuetapowego podejścia do optymalizacji modeli ML oraz przewagę feature engineering w tworzeniu zmiennych o wysokiej wartości predykcyjnej. Permutation importance wykazała, że WorkLife\_Stress\_Score jest najważniejszą cechą dla rzeczywistych predykcji, co potwierdza istotność czynników work-life balance w procesach retencji pracowników.

Praca wnosi znaczący wkład metodologiczny poprzez systematyczne porównanie algorytmów ML w kontekście HR analytics, teoretyczny przez empiryczną weryfikację wpływu feature engineering na wydajność modeli oraz praktyczny przez dostarczenie zoptymalizowanych modeli gotowych do implementacji w systemach HR. Wyniki mają bezpośrednie zastosowanie w organizacjach dążących do data-driven HR management i proactive talent retention strategies.

Słowa kluczowe: machine learning, employee attrition, HR analytics, hyperparameter optimization, feature engineering, predictive modeling, talent retention, XGBoost, regularization.

**1. WPROWADZENIE**

**1.1 Uzasadnienie wyboru tematu**

**1.1.1 Znaczenie problemu rotacji pracowników w współczesnych organizacjach**

W erze gospodarki opartej na wiedzy, gdzie kapitał ludzki stanowi kluczowy czynnik przewagi konkurencyjnej, problem rotacji pracowników (employee attrition) nabiera szczególnego znaczenia. Rotacja pracowników definiowana jako dobrowolne opuszczenie organizacji przez pracownika, stanowi jeden z najkosztowniejszych wyzwań współczesnego zarządzania zasobami ludzkimi.

Dane wykorzystane w niniejszym badaniu, pochodzące z datasetu IBM HR Analytics, wskazują na **16% stopę rotacji** w analizowanej organizacji, co oznacza, że co szósty pracownik opuszcza firmę w ciągu roku. Wskaźnik ten, choć może wydawać się umiarkowany, w kontekście dużej organizacji zatrudniającej tysiące pracowników przekłada się na setki odejść rocznie, generując ogromne koszty i zakłócenia operacyjne.

Podle badań Society for Human Resource Management (SHRM, 2022), średnia stopa rotacji w branży IT i usług technologicznych wynosi 13.2%, podczas gdy w niektórych sektorach może sięgać nawet 25-30%. Oznacza to, że problem rotacji ma charakter systemowy i dotyka większość organizacji, szczególnie w branżach o wysokiej konkurencji o talenty.

**Kluczowe konsekwencje wysokiej rotacji pracowników:**

1. **Bezpośrednie koszty finansowe** - rekrutacja, selekcja, onboarding nowych pracowników
2. **Utrata wiedzy organizacyjnej** - odchodzący pracownicy zabierają ze sobą know-how i doświadczenie
3. **Spadek produktywności** - okresy adaptacji nowych pracowników, obciążenie pozostałego zespołu
4. **Wpływ na morale** - rotacja może mieć efekt domina, wpływając na zadowolenie pozostałych pracowników
5. **Zakłócenia w relacjach z klientami** - szczególnie istotne w organizacjach usługowych

**1.1.2 Koszty rotacji pracowników - analiza ekonomiczna**

Oszacowanie rzeczywistych kosztów rotacji pracowników stanowi complex challenge, wymagający uwzględnienia zarówno kosztów bezpośrednich, jak i pośrednich. Na podstawie analizy literatury przedmiotu oraz benchmarków branżowych, w niniejszym badaniu przyjęto **koszt rotacji jednego pracownika na poziomie 80,000 PLN**.

**Dekompozycja kosztów rotacji:**

**A) Koszty bezpośrednie (40,000 PLN):**

* Proces rekrutacji i selekcji: 15,000 PLN
  + Ogłoszenia rekrutacyjne i portale pracy
  + Czas HR-owców na screening kandydatów
  + Koszty zewnętrznych firm rekrutacyjnych
* Onboarding i szkolenia: 20,000 PLN
  + Programy wprowadzające
  + Szkolenia specjalistyczne
  + Mentoring i supervision
* Koszty administracyjne: 5,000 PLN
  + Rozliczenia z odchodzącym pracownikiem
  + Dokumentacja i procedury

**B) Koszty pośrednie (40,000 PLN):**

* Utrata produktywności: 25,000 PLN
  + Okresy niepełnej wydajności nowego pracownika
  + Dodatkowa praca pozostałego zespołu
  + Możliwe błędy i opóźnienia projektów
* Utrata wiedzy organizacyjnej: 10,000 PLN
  + Know-how specjalistyczne
  + Relacje z klientami i partnerami
  + Nieudokumentowane procesy
* Wpływ na team morale: 5,000 PLN
  + Spadek zaangażowania pozostałych pracowników
  + Potencjalne kolejne odejścia (efekt domina)

Według badań Gallup (2020), koszt zastąpienia jednego pracownika wynosi od 50% do 200% jego rocznego wynagrodzenia, w zależności od poziomu stanowiska. Dla organizacji zatrudniającej 1,000 pracowników z 16% stopą rotacji, roczne koszty rotacji mogą osiągnąć **12.8 miliona PLN** (160 odejść × 80,000 PLN).

**1.1.3 Potencjał technologii i machine learning w HR**

Rozwój technologii analitycznych, w szczególności machine learning, otwiera nowe możliwości w zarządzaniu zasobami ludzkimi. **People Analytics** i **HR Analytics** stanowią dynamicznie rozwijający się obszar, który umożliwia:

1. **Predykcyjne zarządzanie talentami** - przewidywanie problemów zanim wystąpią
2. **Data-driven decision making** - podejmowanie decyzji opartych na danych, a nie intuicji
3. **Personalizacja doświadczeń pracowniczych** - dostosowanie strategii retention do indywidualnych potrzeb
4. **Optymalizacja procesów HR** - automatyzacja i usprawnienie rutynowych działań

**Machine Learning w kontekście attrition prediction oferuje:**

* **Wczesne ostrzeganie** - identyfikacja pracowników zagrożonych odejściem z wyprzedzeniem 3-6 miesięcy
* **Personalizowane interwencje** - dostosowanie działań retention do profilu ryzyka każdego pracownika
* **Continuous monitoring** - stały monitoring wskaźników engagement i satisfaction
* **ROI optimization** - cost-sensitive approach pozwalający na optymalizację zwrotu z inwestycji w retention

**1.1.4 Cost-sensitive learning jako innowacja w HR Analytics**

Tradycyjne podejścia do machine learning w HR koncentrują się na maksymalizacji standardowych metryk jak accuracy czy AUC-ROC. Jednak w kontekście biznesowym, różne typy błędów predykcji mają różne konsekwencje finansowe.

**W przypadku predykcji attrition:**

* **False Negative (FN)** - przeoczenie pracownika, który rzeczywiście odejdzie: **koszt 80,000 PLN**
* **False Positive (FP)** - niepotrzebna interwencja retention: **koszt 3,500 PLN**

Stosunek kosztów FN:FP wynosi około **23:1**, co oznacza, że znacznie bardziej kosztowne jest przeoczenie odejścia niż niepotrzebna interwencja. **Cost-sensitive optimization** pozwala na uwzględnienie tej asymetrii i optymalizację modelu pod kątem minimalizacji całkowitych kosztów biznesowych, a nie tylko poprawy standardowych metryk ML.

Ta innowacyjna metodologia stanowi bridge między world of data science a business requirements, umożliwiając stvarzenie modeli, które nie tylko dobrze przewidują, ale również generują maksymalną wartość biznesową.

**1.2 Problem badawczy i pytania badawcze (2-3 strony)**

**1.2.1 Sformułowanie problemu głównego**

W obliczu rosnących kosztów rotacji pracowników i zwiększającej się konkurencji o talenty, organizacje potrzebują skutecznych narzędzi do predykcji i prewencji odejść pracowników. Tradycyjne metody zarządzania retencją, oparte na periodic surveys i reactive approaches, okazują się niewystarczające w dynamicznym środowisku biznesowym.

**Problem główny niniejszego badania brzmi:**

*"Jak opracować skuteczny model predykcyjny rotacji pracowników wykorzystujący algoritmy uczenia maszynowego, który jednocześnie optymalizuje koszty biznesowe i umożliwia proactive talent retention w organizacji?"*

Problem ten ma charakter wielowymiarowy i obejmuje:

1. **Wymiar techniczny** - wybór i optymalizacja algorytmów ML dla specyfiki danych HR
2. **Wymiar biznesowy** - uwzględnienie rzeczywistych kosztów różnych typów błędów predykcji
3. **Wymiar operacyjny** - zapewnienie implementowalności i użyteczności rozwiązania w praktyce
4. **Wymiar etyczny** - zachowanie fairness i privacy w automatyzacji decyzji HR

**1.2.2 Pytania badawcze szczegółowe**

Dla kompleksowego rozwiązania problemu głównego, sformułowano następujące pytania badawcze szczegółowe:

**PB1: Które algorytmy uczenia maszynowego są najskuteczniejsze w predykcji rotacji pracowników?**

* Jak porównują się modele liniowe vs nieparametryczne w kontekście danych HR?
* Czy modele ensemble przewyższają pojedyncze algorytmy?
* Jaki wpływ ma hyperparameter tuning na skuteczność predykcji?
* Które metryki najlepiej odzwierciedlają jakość modelu w kontekście biznesowym?

**PB2: Jakie czynniki najsilniej wpływają na decyzje o odejściu pracowników?**

* Które zmienne mają największą siłę predykcyjną?
* Jak interpretować wyniki feature importance w kontekście teorii HR?
* Czy istnieją interakcje między zmiennymi zwiększające ryzyko attrition?
* Jakie nowe zmienne można skonstruować przez feature engineering?

**PB3: Jak zoptymalizować model pod kątem kosztów biznesowych?**

* Jaki jest optymalny próg decyzyjny minimalizujący całkowite koszty?
* Jak wpływa cost-sensitive optimization na standardowe metryki ML?
* Jaka jest trade-off między precision a recall w kontekście biznesowym?
* Jak oszacować ROI z implementacji predykcyjnego modelu retention?

**PB4: Jakie są praktyczne implikacje implementacji takiego rozwiązania?**

* Jakie są requirements techniczne i organizacyjne dla wdrożenia?
* Jak zapewnić continuous monitoring i model maintenance?
* Jakie są potencjalne ryzyka i ograniczenia rozwiązania?
* Jak zmierzyć success i business impact po implementacji?

**1.2.3 Kontekst teoretyczny i praktyczny**

Problem rotacji pracowników ma silne osadzenie w teoriach organizational behavior i human resource management. Kluczowe teorie stanowiące foundation dla niniejszego badania to:

**1. Teoria dwuczynnikowa Herzberga (1959)**

* Rozróżnienie między czynnikami higieny (zapobiegającymi niezadowoleniu) a motywatorami (zwiększającymi satysfakcję)
* Implikacje dla identyfikacji key drivers of attrition

**2. Model Job Demands-Resources (JD-R) Bakkera i Demerouti (2007)**

* Równowaga między wymaganiami pracy a dostępnymi zasobami
* Work-life balance jako kluczowy element modelu

**3. Teoria wymiany społecznej (Social Exchange Theory)**

* Reciprocity w relacjach pracodawca-pracownik
* Psychological contract i jego wpływ na commitment

**Praktyczny kontekst** badania obejmuje:

* Rosnące koszty acquisition nowych talentów
* Shortage of skilled workers w branży IT
* Increasing employee expectations dotyczące work-life balance
* Digital transformation w HR processes

**1.3 Cele pracy i hipotezy badawcze (2-3 strony)**

**1.3.1 Cel główny**

**Celem głównym** niniejszego badania jest **opracowanie i walidacja kompleksowego modelu predykcji rotacji pracowników wykorzystującego algoritmy uczenia maszynowego z cost-sensitive optimization, zapewniającego maksymalizację wartości biznesowej przy zachowaniu wysokiej skuteczności predykcyjnej.**

Model ten ma stanowić practical solution dla organizacji dążących do proactive talent management i data-driven HR decisions, oferując konkretne business value w postaci:

* Redukcji kosztów rotacji o minimum 50%
* ROI przekraczający 300% w pierwszym roku
* Skrócenia czasu reakcji na ryzyko attrition z miesięcy do tygodni

**1.3.2 Cele szczegółowe**

**CS1: Analiza porównawcza algorytmów uczenia maszynowego**

* Implementacja i ewaluacja 5 różnych algorytmów ML (Logistic Regression, SVM, Random Forest, XGBoost, Extra Trees)
* Comprehensive hyperparameter tuning dla każdego algorytmu
* Statistical comparison wyników z wykorzystaniem cross-validation
* Identyfikacja najskuteczniejszego algorytmu dla danych HR

**CS2: Identyfikacja i interpretacja kluczowych predyktorów attrition**

* Feature importance analysis z wykorzystaniem multiple methods
* Business interpretation wyników w kontekście teorii HR
* Categorization predyktorów według obszarów wpływu (compensation, work-life balance, career development, etc.)
* Development nowych zmiennych przez advanced feature engineering

**CS3: Implementacja cost-sensitive optimization**

* Quantification rzeczywistych kosztów biznesowych błędów predykcji
* Development cost-sensitive threshold optimization
* ROI analysis i business case development
* Comparison standardowych metryk ML vs business-oriented metrics

**CS4: Validacja praktycznej użyteczności rozwiązania**

* Assessment implementability w real-world environment
* Development practical recommendations dla HR professionals
* Risk assessment i mitigation strategies
* Framework dla continuous model monitoring i improvement

**1.3.3 Hipotezy badawcze**

Na podstawie przeglądu literatury i preliminary analysis danych, sformułowano następujące hipotezy badawcze:

**H1: Algorytmy ensemble przewyższają modele liniowe w predykcji attrition**

*Uzasadnienie:* Dane HR charakteryzują się complex non-linear relationships i interactions między zmiennymi. Modele ensemble jak Random Forest czy XGBoost powinny lepiej capture te zależności niż linear models.

*Operacjonalizacja:* AUC-ROC modeli ensemble > AUC-ROC Logistic Regression o minimum 0.05

**H2: Feature engineering znacząco poprawia skuteczność predykcji**

*Uzasadnienie:* Raw variables w datasecie HR często nie capture pełnego business context. Engineering nowych zmiennych opartych na domain knowledge powinien improve model performance.

*Operacjonalizacja:* AUC-ROC modelu z engineered features > AUC-ROC modelu z raw features o minimum 0.08

**H3: Cost-sensitive optimization generuje ROI przekraczający 500%**

*Uzasadnienie:* Traditional ML metrics nie uwzględniają business costs różnych typów błędów. Cost-sensitive approach powinien significantly improve business value przy marginalnym spadku technical metrics.

*Operacjonalizacja:* ROI z cost-optimized model ≥ 500% w pierwszym roku operacyjnym

**H4: Work-life balance stanowi najważniejszy predyktor attrition**

*Uzasadnienie:* Literature review i preliminary data analysis wskazują na overriding importance czynników related do work-life balance, szczególnie overtime i job satisfaction.

*Operacjonalizacja:* Variables related do work-life balance (OverTime, WorkLifeBalance, JobSatisfaction) stanowią ≥40% total feature importance

**1.3.4 Oczekiwany wkład do wiedzy**

**Wkład metodologiczny:**

* Development integrated framework łączącego ML techniques z business optimization
* Novel approach do cost-sensitive learning w HR context
* Practical methodology dla ROI assessment ML solutions w HR

**Wkład teoretyczny:**

* Empirical validation key theories of employee retention
* Quantitative analysis względnej ważności różnych retention factors
* Evidence-based insights dla HR theory development

**Wkład praktyczny:**

* Ready-to-implement solution dla HR professionals
* Actionable recommendations oparte na data-driven insights
* Template dla scaling podobnych solutions w różnych organizacjach

**2. PRZEGLĄD LITERATURY**

**2.1 Teoretyczne podstawy rotacji pracowników**

Rotacja pracowników jest zjawiskiem złożonym, leżącym na pograniczu psychologii pracy, socjologii organizacji oraz ekonomii. Zrozumienie mechanizmów leżących u podstaw decyzji o opuszczeniu organizacji wymaga odwołania się do ugruntowanych **teorii motywacji i zachowań organizacyjnych**. Te ramy teoretyczne nie tylko wyjaśniają, dlaczego pracownicy odchodzą, ale także stanowią **podstawę doboru i inżynierii cech predykcyjnych** w modelach uczenia maszynowego. Szczegółowy przegląd literatury koncentruje się na trzech kluczowych perspektywach wyjaśniających proces rotacji: teorii dwuczynnikowej Herzberga, modelu Job Demands-Resources (JD-R) oraz teorii wymiany społecznej.

**2.1.1 Teoria dwuczynnikowa Herzberga**

Teoria dwuczynnikowa, zwana również teorią motywacyjno-higieniczną, została sformułowana przez Fredericka Herzberga w 1959 roku. Wprowadza ona dychotomiczny podział czynników wpływających na **zadowolenie i niezadowolenie z pracy**. W kontekście rotacji, teoria ta dostarcza istotnych wniosków dotyczących tego, które aspekty środowiska pracy mają potencjał retencyjny, a które jedynie zapobiegają niezadowoleniu.

Herzberg wyróżnił dwie niezależne kategorie czynników:

1. **Czynniki higieny (*Hygiene Factors*)**: Odpowiadają za **zapobieganie niezadowoleniu** (a zatem pośrednio za redukcję *attrition*). Związane są z kontekstem pracy i obejmują: politykę firmy i administrację, wynagrodzenie, bezpieczeństwo zatrudnienia, warunki pracy, relacje interpersonalne oraz nadzór. Ich brak lub nieodpowiedni poziom prowadzi do **niezadowolenia i potencjalnej rotacji**, jednak ich obecność (nawet na wysokim poziomie) **nie generuje trwałej motywacji** ani głębokiego zaangażowania.
2. **Czynniki motywacyjne (*Motivators*)**: Odpowiadają za **wzrost zadowolenia i motywacji**. Związane są z treścią pracy i osiągnięciami, w tym: osiągnięcia, uznanie, charakter samej pracy, odpowiedzialność, awans i możliwości rozwoju. To właśnie te czynniki mają **potencjał retencyjny** w dłuższej perspektywie, ponieważ wpływają na wewnętrzną satysfakcję pracownika.

**Implikacje dla predykcji ML:** Teoria Herzberga sugeruje, że model predykcyjny powinien uwzględniać zarówno **zmienne higieniczne** (np. poziom wynagrodzenia, relacje z menedżerem, dojazdy) jako wskaźniki ryzyka rotacji, jak i **zmienne motywacyjne** (np. możliwości awansu, satysfakcja z pracy, poziom odpowiedzialności) jako wskaźniki retencji. Wysoka wartość czynników higieny może obniżyć ryzyko, ale dopiero wysoka wartość czynników motywacyjnych je stabilizuje.

**2.1.2 Model Job Demands-Resources (JD-R)**

Model Obciążenia-Zasobów Pracy (*Job Demands-Resources Model, JD-R*), sformułowany przez Demerouti, Bakker, Nachreiner i Schaufeli (2001), jest współczesnym, elastycznym modelem psychologii pracy. Wyjaśnia on, jak interakcja dwóch kluczowych kategorii czynników środowiska pracy wpływa na **wypalenie zawodowe, zaangażowanie (engagement)**, a w konsekwencji na decyzję o rotacji.

JD-R opiera się na założeniu dwóch odrębnych procesów psychologicznych:

1. **Proces obciążenia zdrowotnego (*Health Impairment Process*)**: Wywołany przez **wymagania/obciążenia pracy (*Job Demands*)**, które odnoszą się do fizycznych, psychologicznych, społecznych lub organizacyjnych aspektów pracy, wymagających stałego wysiłku. Przykłady obejmują: nadmierną presję czasową, dużą odpowiedzialność, konflikty ról. Wysokie i chroniczne obciążenia prowadzą do **wypalenia zawodowego** i negatywnie korelują ze zdrowiem i samopoczuciem, co stanowi **silny predyktor odejścia**.
2. **Proces motywacyjny (*Motivational Process*)**: Wywołany przez **zasoby pracy (*Job Resources*)**, które stanowią fizyczne, psychologiczne, społeczne lub organizacyjne aspekty pracy, pomagające w osiąganiu celów, redukcji obciążeń oraz stymulowaniu rozwoju. Przykłady obejmują: autonomię, wsparcie społeczne, *feedback*, szkolenia, możliwości awansu. Zasoby te prowadzą do **zaangażowania w pracę** (*work engagement*), które jest **kluczowym czynnikiem retencyjnym**.

**Implikacje dla predykcji ML:** Model JD-R uzasadnia tworzenie **zmiennych kompozytowych** (jak zaproponowany WorkLife\_Stress\_Score), które agregują różne obciążenia. Podkreśla on znaczenie **interakcji** między Obciążeniami a Zasobami: zasoby mogą buforować negatywny wpływ wysokich obciążeń. W modelowaniu predykcyjnym, niska wartość zasobów (np. brak *feedbacku*, brak szkoleń) w połączeniu z wysokimi obciążeniami jest sygnałem **najwyższego ryzyka rotacji**.

**2.1.3 Teoria wymiany społecznej**

Teoria wymiany społecznej (*Social Exchange Theory, SET*), zapoczątkowana przez Homansa (1961) i Blau (1964), jest paradygmatem socjologicznym, który analizuje relacje międzyludzkie jako serię **transakcji opartych na wzajemności**. W kontekście organizacji, teoria ta wyjaśnia, że pracownicy oceniają swoją relację z pracodawcą w kategoriach **wkładów (*inputs*)** i **otrzymanych korzyści (*outcomes*)**.

Kluczowe założenie SET w HR polega na istnieniu **kontraktu psychologicznego** (*psychological contract*) – niezapisanego zestawu wzajemnych oczekiwań. Jeśli pracownik postrzega relację jako sprawiedliwą i **wzajemną** (wkłady = korzyści), rośnie jego poczucie **zobowiązania organizacyjnego (*organizational commitment*)** i skłonność do pozostania.

**Mechanizmy rotacji według SET:**

1. **Negatywne postrzeganie wzajemności**: Pracownik uważa, że jego wkład (wysiłek, godziny, *know-how*) jest **niedostatecznie wynagradzany** przez korzyści (wynagrodzenie, uznanie, rozwój, bezpieczeństwo).
2. **Percepcja niesprawiedliwości (Equity Theory)**: Niesprawiedliwość w traktowaniu w porównaniu z innymi pracownikami (*Deviations analysis*).
3. **Wycofanie się**: Niewywiązywanie się pracodawcy z kontraktu psychologicznego prowadzi do obniżenia zaangażowania, *job embeddedness* (zakorzenienia w pracy), a ostatecznie do **wycofania się z relacji** (rotacji).

**Implikacje dla predykcji ML:** SET dostarcza uzasadnienia dla zmiennych mierzących **sprawiedliwość i równowagę** relacji. W szczególności, uzasadnia zastosowanie **analizy dewiacji** (np. MonthlyIncome\_Deviation\_from\_Department), która mierzy postrzeganie sprawiedliwości wynagrodzenia w odniesieniu do grupy referencyjnej. Niska percepcja sprawiedliwości traktowania (niski wynik na osi *outcome*) przy wysokim zaangażowaniu (wysoki *input*) jest silnym sygnałem **dysfunkcyjnej relacji** i ryzyka rotacji.

**2.2 HR Analytics i People Analytics**

Rozwój technologii informatycznych, wzrost mocy obliczeniowej oraz dostępność dużych zbiorów danych (Big Data) doprowadziły do transformacji w zarządzaniu zasobami ludzkimi (ZZL). Kluczową rolę w tej przemianie odgrywają **HR Analytics** i **People Analytics**. Choć terminy te są często używane zamiennie, ogólnie przyjmuje się, że **HR Analytics** koncentruje się na mierzeniu efektywności procesów HR (np. czas rekrutacji, koszt szkolenia), natomiast **People Analytics** jest szerszym i bardziej zaawansowanym paradygmatem, wykorzystującym metody statystyczne i **uczenie maszynowe (ML)** do **predykcji zachowań** i optymalizacji decyzji biznesowych związanych z kapitałem ludzkim (Marler & Boudreau, 2017).

**2.2.1 Ewolucja analityki w zarządzaniu zasobami ludzkimi**

Ewolucję analityki w ZZL można podzielić na cztery główne etapy. Każdy kolejny etap zwiększa potencjał do generowania strategicznej wartości biznesowej, przechodząc od **opisu** do **predykcji i preskrypcji**.

| Etap | Typ Analityki | Cel | Pytanie, na które odpowiada | Związek z Rotacją |
| --- | --- | --- | --- | --- |
| **I** | **Opisowa (*Descriptive*)** | Raportowanie i podsumowanie danych historycznych. | Co się stało? | Obliczanie historycznej stopy rotacji (np. 16,1%). |
| **II** | **Diagnostyczna (*Diagnostic*)** | Identyfikacja przyczyn zdarzeń. | Dlaczego to się stało? | Ustalenie, że rotacja jest wyższa w dziale X lub wśród pracowników z niskim wynagrodzeniem. |
| **III** | **Predykcyjna (*Predictive*)** | Prognozowanie przyszłych zdarzeń i zachowań. | Co się stanie? | **Identyfikacja pracowników zagrożonych odejściem** z 3-6-miesięcznym wyprzedzeniem – kluczowy cel niniejszej pracy. |
| **IV** | **Preskryptywna (*Prescriptive*)** | Dostarczanie rekomendacji do optymalnych działań. | Co należy zrobić? | Zasugerowanie optymalnej interwencji (np. podwyżka vs. awans) dla zminimalizowania kosztów rotacji. |

Eksportuj do Arkuszy

Obecny etap rozwoju People Analytics, na którym koncentruje się niniejsza praca, to etap III i IV. **Analityka predykcyjna** (etap III), wykorzystująca algorytmy ML, pozwala na transformację strategii HR z **reaktywnej** na **proaktywną**. Modele uczenia maszynowego (np. Regresja Logistyczna, Random Forest, Gradient Boosting) są trenowane na danych historycznych, aby określić **prawdopodobieństwo rotacji** dla każdego pracownika, umożliwiając interwencje *just-in-time* i optymalizację kosztów, zgodnie z celami badawczymi.

**2.2.2 Zastosowania ML w Talent Management**

Uczenie maszynowe jest obecnie fundamentalnym narzędziem w zarządzaniu talentami (*Talent Management*), wykraczającym poza prostą predykcję rotacji. Jego zastosowanie jest widoczne w kluczowych filarach ZZL:

1. **Predykcja Rotacji (*Attrition Prediction*)**: Najbardziej dojrzałe zastosowanie. Polega na klasyfikacji pracowników jako „wysokie ryzyko odejścia” lub „niskie ryzyko odejścia”. Zgodnie z przeglądem literatury (np. Turetski, 2017), modele drzewiaste (np. **Random Forest, XGBoost**) często wykazują lepszą wydajność niż tradycyjne modele statystyczne, szczególnie w przypadku skomplikowanych danych nieliniowych, typowych dla HR. Modele te nie tylko przewidują, *kto* odejdzie, ale również identyfikują **najważniejsze czynniki ryzyka (*Feature Importance*)**, co jest niezbędne do projektowania efektywnych interwencji retencyjnych.
2. **Rekrutacja i Selekcja**: ML służy do optymalizacji źródeł kandydatów, automatycznego *screeningu* CV oraz predykcji **dopasowania kulturowego (*Cultural Fit*)** i **długoterminowej wydajności** kandydatów (np. za pomocą przetwarzania języka naturalnego – NLP).
3. **Zarządzanie Wydajnością (*Performance Management*)**: Modele predykcyjne mogą identyfikować pracowników z wysokim potencjałem (*High Potentials*) oraz tych, którzy są zagrożeni spadkiem wydajności, co pozwala na proaktywne planowanie szkoleń i *coaching*.
4. **Optymalizacja Ścieżek Kariery i Szkoleń**: Wykorzystanie ML do mapowania optymalnych ścieżek kariery i rekomendowania spersonalizowanych szkoleń (*Personalized Learning*) w celu maksymalizacji retencji i rozwoju.

Niniejsza praca koncentruje się na zastosowaniu zaawansowanej metodyki ML w predykcji rotacji w celu **minimalizacji kosztów biznesowych** związanych z niekontrolowanymi odejściami.

**2.2.3 Etyczne aspekty AI w HR**

Wzrost wykorzystania sztucznej inteligencji (AI) i uczenia maszynowego w procesach decyzyjnych HR rodzi istotne **wyzwania etyczne i prawne**. Aby zachować zgodność z regulacjami (np. RODO/GDPR) i standardami etycznymi, niezbędna jest świadomość potencjalnych pułapek.

1. **Dyskryminacja Algorytmiczna i Uprzedzenia (*Bias*)**: Modele ML trenowane są na historycznych danych. Jeśli dane te odzwierciedlają historyczne uprzedzenia (np. niższe wskaźniki awansów dla określonych grup demograficznych), model może **utrwalić i wzmocnić dyskryminację** (np. w predykcji awansów lub rotacji). Wymaga to metodyki **sprawiedliwości algorytmicznej (*Algorithmic Fairness*)** i rygorystycznego testowania modelu pod kątem uprzedzeń.
2. **Przejrzystość (*Transparency*) i Wyjaśnialność (*Explainability*)**: W wielu jurysdykcjach pracownicy mają prawo do zrozumienia, **dlaczego** algorytm podjął określoną decyzję (np. zakwalifikował ich jako „wysokie ryzyko rotacji”). Konieczne jest stosowanie metod **XAI (*Explainable AI*)**, takich jak wartości **SHAP** lub **LIME**, które zwiększają interpretowalność wyników nawet dla złożonych modeli *black box* (np. XGBoost).
3. **Prywatność Danych (*Data Privacy*)**: Gromadzenie i przetwarzanie wrażliwych danych pracowniczych (np. dane demograficzne, monitoring aktywności) wymaga ścisłej zgodności z przepisami o ochronie danych. W kontekście rotacji predykcyjnej, istotne jest **anonimizowanie** lub **pseudonimizowanie** zbioru danych oraz zapewnienie minimalizacji danych (*data minimization*).

W niniejszym badaniu, aspekt etyczny jest uwzględniony poprzez użycie **publicznego, zanonimizowanego zbioru danych** oraz przez skupienie się na **wyjaśnialności modelu** poprzez analizę istotności cech, co pozwala na transparentne uzasadnienie wyników i uniknięcie bezkrytycznego polegania na prognozach „czarnej skrzynki”.

Rozumiem. Aby sekcja była kompletna, dodam szczegółowe uzasadnienie wyboru algorytmu **Extra Trees** do modeli *ensemble*.

Poniżej znajduje się pełna propozycja treści dla pozostałej części rozdziału 2., z uwzględnieniem dodatkowego algorytmu.

**2.3 Algorytmy uczenia maszynowego w klasyfikacji**

W predykcji rotacji pracowników, kluczowym zadaniem jest **klasyfikacja binarna** (pracownik *odejdzie* vs. pracownik *pozostanie*). Wybór odpowiednich algorytmów ma fundamentalne znaczenie dla uzyskania wysokiej **dokładności predykcyjnej** oraz **wyjaśnialności modelu**. Niniejsze badanie uwzględnia różnorodne rodziny algorytmów, aby zrównoważyć transparentność z wydajnością.

**2.3.1 Modele liniowe: Regresja Logistyczna (*Logistic Regression*)**

**Regresja Logistyczna (LR)** jest jednym z najczęściej stosowanych modeli w analizie predykcyjnej w HR. Jest to model liniowy, który wykorzystuje funkcję logit do oszacowania prawdopodobieństwa przynależności obserwacji do jednej z dwóch klas, opierając się na regresji liniowej cech wejściowych.

**Uzasadnienie Wyboru do Predykcji Rotacji:**

1. **Interpretowalność (*Explainability*)**: Regresja Logistyczna jest modelem **wysoce transparentnym (biała skrzynka)**. Współczynniki modelu ((βi​)) bezpośrednio wskazują kierunek i siłę wpływu każdej cechy (np. wynagrodzenia, stażu pracy) na ryzyko rotacji. Jest to kluczowe w HR, gdzie menedżerowie potrzebują **wyjaśnień decyzyjnych**, a nie tylko prognoz.
2. **Punkt Odkłócenia (*Baseline*)**: Stanowi solidny punkt odniesienia (*baseline*) dla oceny bardziej złożonych algorytmów, pozwalając na weryfikację, czy zysk w dokładności predykcyjnej modeli *ensemble* jest wart utraty części interpretowalności.

**2.3.2 Modele *ensemble*: Random Forest, Extra Trees, XGBoost**

Modele *ensemble* (zespołowe) łączą predykcje wielu pojedynczych modeli (*weak learners*), najczęściej drzew decyzyjnych, w celu uzyskania znacznie lepszej i bardziej stabilnej wydajności niż pojedynczy klasyfikator. W predykcji rotacji są one wybierane ze względu na zdolność do modelowania **nieliniowych i interakcyjnych relacji** między zmiennymi.

**A) Las Losowy (*Random Forest*, RF)**

**Random Forest** jest algorytmem typu **bagging** (Bootstrap Aggregating), który buduje dużą liczbę nieskorelowanych drzew decyzyjnych. Każde drzewo jest trenowane na losowej próbce danych i losowym podzbiorze cech. Ostateczna predykcja jest wynikiem głosowania większości. Jest to algorytm ceniony za **odporność na przeuczenie (*overfitting*)** i efektywność w identyfikacji ważności cech.

**B) Ekstremalnie Losowe Drzewa (*Extra Trees*, ET)**

**Extra Trees** (Extremely Randomized Trees) to rozszerzenie idei Random Forest. Choć również opiera się na budowaniu wielu drzew na losowych podpróbkach, wprowadza **dodatkowy poziom losowości** w procesie podziału. Zamiast szukać optymalnego progu podziału dla każdej cechy, jak robi to RF, ET wybiera progi podziału **całkowicie losowo**.

**Uzasadnienie Wyboru:** Model ten charakteryzuje się zazwyczaj **szybszym czasem uczenia** (ponieważ pomija kosztowny etap optymalizacji progu podziału) oraz dalszą **redukcją wariancji** w porównaniu do RF. Włączając ET do porównania, możliwe jest zbadanie, jak zwiększona losowość wpływa na zdolność modelu do generalizacji na nieznane dane i czy nieznacznie niższy potencjalny *bias* przewyższa utratę precyzji w podziałach.

**C) Ekstremalny Gradient Boosting (*XGBoost*)**

**XGBoost** (eXtreme Gradient Boosting) jest algorytmem typu **boosting**, który sekwencyjnie buduje nowe drzewa decyzyjne, koncentrując się na korygowaniu błędów popełnionych przez poprzednie drzewa. Uważany jest za jeden z **najbardziej wydajnych** i skalowalnych algorytmów klasyfikacji.

**Uzasadnienie Wyboru:** XGBoost często osiąga **najwyższe metryki wydajności** (*performance*) w zadaniach klasyfikacyjnych. Jego zdolność do uchwycenia bardzo złożonych, **nieliniowych interakcji** między zmiennymi jest kluczowa przy modelowaniu subtelnych mechanizmów rotacji. Włączenie tego algorytmu ma na celu osiągnięcie **maksymalnego potencjału predykcyjnego** modelu.

**2.3.3 Modele nieparametryczne: SVM, KNN**

Wybór modeli nieparametrycznych w początkowej fazie badania pozwala na ocenę struktury danych bez przyjmowania z góry określonych założeń dystrybucyjnych.

**A) Metoda Wektorów Nośnych (*Support Vector Machine, SVM*)**

**SVM** jest algorytmem, którego celem jest znalezienie **optymalnej hiperpłaszczyzny** maksymalizującej margines między dwiema klasami w przestrzeni cech. W przypadku danych nieliniowych, SVM używa **jąder (*kernels*)** do mapowania danych do wyższego wymiaru, co umożliwia skuteczną klasyfikację.

**Uzasadnienie Wyboru:** SVM jest przydatny do sprawdzenia, czy granica decyzyjna między pracownikami odchodzącymi a pozostającymi ma **wyraźnie nieliniowy charakter**, co nie jest możliwe do zbadania przy użyciu modeli liniowych.

**B) K-Najbliższych Sąsiadów (*K-Nearest Neighbors, KNN*)**

**KNN** jest prostym, opartym na instancjach algorytmem, który klasyfikuje nowy punkt danych na podstawie większości głosów jego **K najbliższych sąsiadów** w przestrzeni cech, mierzonej za pomocą metryki odległości.

**Uzasadnienie Wyboru:** KNN dostarcza wglądu w **lokalne struktury danych**. Jeśli pracownicy odchodzący i pozostający tworzą **wyraźne, odrębne klastry** w przestrzeni cech oparte na bliskości (podobieństwie profilu), KNN może to skutecznie zidentyfikować, służąc jako prosta, ale mocna kontrola dla innych, bardziej złożonych modeli.

**2.4 Cost-sensitive learning i optymalizacja biznesowa**

Tradycyjne podejście do uczenia maszynowego koncentruje się na maksymalizacji **dokładności predykcyjnej** (*Accuracy*) lub ogólnej miary (*F1-score*). Jednakże w kontekście biznesowym i problemu rotacji pracowników, **koszty błędnej klasyfikacji nie są symetryczne**.

1. **Błąd I typu (*False Positive*)**: Model przewidział rotację, ale pracownik **pozostał**. Koszt: niepotrzebne wdrożenie kosztownej interwencji retencyjnej (np. podwyżka, szkolenie).
2. **Błąd II typu (*False Negative*)**: Model przewidział pozostanie, ale pracownik **odszedł**. Koszt: **realne i wysokie koszty rotacji** (rekrutacja, utrata *know-how*).

Powszechnie przyjmuje się, że koszt Błędu II typu jest **wielokrotnie wyższy** niż koszt Błędu I typu. Celem niniejszej pracy jest więc zastosowanie metodyki **Cost-Sensitive Learning**, co wymaga:

* **Optymalizacji metryki *Recall***: Koncentracja na minimalizacji Błędów II typu (fałszywie negatywnych), co oznacza maksymalizację wykrywalności pracowników *faktycznie* zagrożonych odejściem.
* **Wprowadzenie Macierzy Kosztów Biznesowych**: Zamiast polegania wyłącznie na metrykach statystycznych, wykorzystanie **rzeczywistej macierzy kosztów**, która kwantyfikuje straty finansowe związane z każdym typem błędu.
* **Optymalizacja Finansowa**: Tunowanie hiperparametrów nie pod kątem najwyższego F1-score, lecz pod kątem **maksymalizacji zysku netto** lub minimalizacji straty operacyjnej dla organizacji.

To podejście odróżnia niniejszą pracę od typowych akademickich studiów przypadku i jest zgodne z jej głównym celem: **optymalizacją kosztów biznesowych**.

**2.5 Identyfikacja luki badawczej**

Na podstawie szczegółowego przeglądu literatury, niniejsze badanie identyfikuje i dąży do wypełnienia następującej **luki badawczej**:

Istniejące prace dotyczące predykcji rotacji z wykorzystaniem uczenia maszynowego w HR (rozdział 2.2) często koncentrują się na maksymalizacji metryk statystycznych, takich jak *Accuracy* lub *F1-score*, z pominięciem lub marginalizacją **rzeczywistego kosztu biznesowego** błędnej klasyfikacji (rozdział 2.4). Ponadto, choć teorie rotacji (rozdział 2.1) oferują bogate ramy koncepcyjne, niewiele badań systematycznie przekłada wnioski z **Teorii Dwuczynnikowej, JD-R i SET** na innowacyjną i kompleksową metodologię **Feature Engineering** (rozdział 1.1.4).

**Wkład Niniejszej Pracy:**

Niniejsza praca wypełnia lukę poprzez:

1. **Integrację Teoretyczną i Metodyczną**: Systematyczne przełożenie założeń teorii motywacji (JD-R, SET) na **zaawansowane, kompozytowe cechy predykcyjne**.
2. **Optymalizację Zorientowaną na Koszty**: Zastosowanie **Cost-Sensitive Learning** i walidację modeli na podstawie **maksymalizacji finansowego zysku netto** dla organizacji, a nie tylko metryk statystycznych.
3. **Porównawczą Analizę Wyjaśnialności**: Przeprowadzenie analizy wydajności i transparentności modeli z różnych rodzin algorytmów (liniowe, *ensemble*), co jest krytyczne dla praktycznego wdrożenia wyników w działach HR.

W ten sposób badanie ma ambicję dostarczenia **praktycznego, finansowo zorientowanego i solidnego** modelu do zarządzania rotacją pracowników.