

Technika	Opis	Zalety	Wady	Przykłady zastosowań	Przykład
One-Hot Encoding	Prosta binarna reprezentacja kategorii.	Łatwość implementacji.	Wysoka wymiarowość, brak relacji między kategoriami.	Klasyfikacja prostych danych kategoriycznych.	"kot" → [1, 0, 0], "pies" → [0, 1, 0].
Word2Vec	Embedding słów uczony w kontekście (CBOW, Skip-gram).	Zachowuje semantykę i syntaktykę słów.	Wektory są statyczne, brak uwzględnienia kontekstu zdania.	Przetwarzanie języka naturalnego (NLP).	"król" - "mężczyzna" + "kobieta" = "królowa".
GloVe	Globalna analiza współwystępowania słów w korpusie.	Zachowuje globalne zależności słów.	Styczne embeddingi, brak dynamicznego kontekstu.	NLP, analiza semantyczna tekstów.	"samochód" ≈ "pojazd" w przestrzeni wektorowej.
FastText	Uwzględnia subwordy, co poprawia działanie dla słów rzadkich lub nowych.	Obsługuje słowa spoza korpusu.	Wolniejszy trening niż Word2Vec.	NLP, szczególnie języki złożone morfologicznie.	"piesek" ≈ "pies" dzięki wspólnym subwordom.
BERT, GPT, RoBERTa	Kontekstowe modele językowe, embeddingi zależne od kontekstu zdania.	Dynamiczne reprezentacje słów.	Wysokie wymagania obliczeniowe.	Tłumaczenia maszynowe, chatboty, analiza tekstów.	"bank" (rzeka) ≠ "bank" (finanse) w różnych kontekstach.
Warstwy embedding w NN	Embedding generowany podczas treningu sieci neuronowej.	Automatyczna optymalizacja reprezentacji danych.	Wymaga odpowiednich danych i treningu.	Rekomendacje, klasyfikacja, NLP.	Filmy i użytkownicy w systemie rekomendacji Netflix.
Node2Vec	Embedding węzłów w grafie z losowymi przejściami (jak Word2Vec).	Zachowuje strukturalne podobieństwa.	Ograniczone do grafów.	Analiza sieci społecznościowych, grafów wiedzy.	"Użytkownik A" ≈ "Użytkownik B" w grafie znajomości.
DeepWalk	Technika losowych przejść w grafie z embeddingami jak Word2Vec.	Działa na dużych grafach.	Brak uwzględnienia szczegółowej struktury grafu.	Analiza powiązań w grafach.	Klaster użytkowników z podobnymi preferencjami.
Graph Neural Networks (GNNs)	Sieci neuronowe analizujące strukturę grafu.	Uczy się zależności między węzłami.	Wymaga dużej mocy obliczeniowej.	Rekomendacje, analiza biologiczna, sieci wiedzy.	Relacje między genami w badaniach biologicznych.
PCA (Principal Component Analysis)	Redukcja wymiarów z zachowaniem kluczowych cech danych.	Prostota, szybka analiza.	Może tracić ważne informacje.	Wizualizacja i analiza danych liczbowych.	Dane 5D zredukowane do 2D dla wizualizacji.
t-SNE	Wizualizacja danych w niskiej wymiarowości.	Intuicyjne klastrowanie.	Nieefektywne dla dużych zbiorów danych.	Wizualizacja klastrów.	Klastry użytkowników według ich preferencji.
UMAP	Podobne do t-SNE, ale szybsze i bardziej dokładne.	Lepsze odzwierciedlenie struktury danych.	Może wymagać dostrajania parametrów.	Wizualizacja, analiza klastrów danych.	Wizualizacja klastrów dokumentów w 2D.
Autoenkodery	Sieci neuronowe kompresujące dane do niskiej wymiarowości.	Wydajne dla złożonych danych, np. obrazy.	Wymaga dużej mocy obliczeniowej, trudne w treningu.	Obrazy, dźwięk, dane czasowe.	Obraz twarzy zredukowany do kluczowych cech.
Matrix Factorization	Faktoryzacja macierzy współzależności, np. w systemach rekomendacji.	Prosta implementacja, skuteczność w rekomendacjach.	Traci kontekst, nie działa dla nowego użytkownika/obiektu.	Systemy rekomendacji (Netflix, Amazon).	Użytkownik ≈ Produkt na podstawie współzależności.
Predefiniowane embeddingi	Wykorzystanie gotowych modeli, np. GloVe, Word2Vec, BERT.	Szybkie wdrożenie, brak potrzeby trenowania od zera.	Mogą być niedostosowane do specyficznych danych użytkownika.	NLP, szybkie prototypowanie modeli.	Słowo "data" → embedding BERT.