

Spis treści |

Słowo wstępne	13
O autorach	15
O recenzentach	16
Przedmowa	17

ROZDZIAŁ 1

Nawigowanie po krajobrazie NLP — kompleksowe wprowadzenie 22

Dla kogo jest ta książka?	23
Co to jest przetwarzanie języka naturalnego?	23
Historia i ewolucja przetwarzania języka naturalnego	23
Wstępne strategie maszynowego przetwarzania języka naturalnego	24
Zwycięska synergia — połączenie NLP i ML	26
Wprowadzenie do matematyki i statystyki w NLP	27
Przykład modelu językowego — ChatGPT	30
Podsumowanie	30
Pytania i odpowiedzi	31

ROZDZIAŁ 2

Algebra liniowa, prawdopodobieństwo i statystyka

w uczeniu maszynowym i NLP 32

Wprowadzenie do algebry liniowej	32
Podstawowe działania na macierzach i wektorach	35
Definicje macierzy	36
Wartości i wektory własne	37
Metody numeryczne znajdowania wektorów własnych	38
Rozkład macierzy na wartości własne	38
Rozkład według wartości osobliwych	39
Prawdopodobieństwo w uczeniu maszynowym	40
Niezależność statystyczna	41
Zmienne losowe dyskretne i ich rozkład	42

Funkcja gęstości prawdopodobieństwa	42
Estymacja bayesowska	48
Podsumowanie	49
Dalsza lektura	49
Bibliografia	50

ROZDZIAŁ 3

Wykorzystanie potencjału uczenia maszynowego w NLP	52
Wymagania techniczne	52
Eksploatacja danych	53
Wizualizacja danych	54
Oczyszczanie danych	55
Selekcja cech	59
Inżynieria cech	68
Typowe modele uczenia maszynowego	73
Regresja liniowa	73
Regresja logistyczna	74
Drzewa decyzyjne	75
Las losowy	77
Maszyny wektorów nośnych (SVM)	79
Sieci neuronowe i transformery	80
Niedostateczne i nadmierne dopasowanie modelu	82
Dzielenie danych	88
Dostrajanie hiperparametrów	89
Modele zespołowe	91
Bagging	91
Wzmacnianie	92
Spiętrzanie	94
Lasy losowe	95
Wzmacnianie gradientowe	95
Dane niezrównoważone	96
SMOTE	97
Algorytm NearMiss	98
Uczenie wrażliwe na koszty	99
Wzbogacanie danych	100
Dane skorelowane	101
Podsumowanie	102
Bibliografia	102

ROZDZIAŁ 4**Usprawnianie technik wstępnego przetwarzania tekstu**

pod kątem optymalnej wydajności NLP	103
Wymagania techniczne	104
Normalizacja tekstu	104
Zamiana na małe litery	104
Usuwanie znaków specjalnych i interpunkcyjnych	104
Usuwanie słów stopu	105
Sprawdzanie i poprawianie pisowni	106
Lematyzacja	106
Tematyzacja	106
Rozpoznawanie nazwanych encji (NER)	107
Oznaczanie części mowy	109
Metody oparte na regułach	110
Metody statystyczne	110
Metody oparte na uczeniu głębokim	111
Wyrażenia regularne	112
Tokenizacja	116
Potok wstępnego przetwarzania tekstu	117
Kod NER i POS	118
Podsumowanie	119

ROZDZIAŁ 5**Klasyfikowanie tekstu — wykorzystanie tradycyjnych technik**

uczenia maszynowego	120
Wymagania techniczne	121
Typy klasyfikacji tekstu	121
Uczenie nadzorowane	122
Uczenie nienadzorowane	124
Uczenie półnadzorowane	125
Klasyfikacja zdań z wykorzystaniem reprezentacji wektorowej z kodowaniem z „gorącą jedyneką”	126
Klasyfikacja tekstu metodą TF-IDF	129
Klasyfikacja tekstu z użyciem Word2Vec	132
Word2Vec	132
Ewaluacja modelu	135
Nadmierne i niedostateczne dopasowanie	137
Dostrajanie hiperparametrów	139
Dodatkowe zagadnienia związane z praktyczną klasyfikacją tekstu	140

Modelowanie tematyczne — praktyczne zastosowanie	
nienadzorowanej klasyfikacji tekstu	141
LDA	142
Projekt rzeczywistego systemu ML do klasyfikacji tekstu	144
Implementowanie rozwiązania ML	147
Przykładowy scenariusz — projekt systemu ML do klasyfikacji NLP	
w notatniku Jupytera	149
Cel biznesowy	149
Cel techniczny	149
Potok	149
Ustawienia	150
Selekcja cech	154
Generowanie wybranego modelu	154
Podsumowanie	155

ROZDZIAŁ 6

Nowe spojrzenie na klasyfikowanie tekstu — językowe modele uczenia głębokiego

157

Wymagania techniczne	158
Podstawy uczenia głębokiego	159
Co to jest sieć neuronowa?	159
Podstawowa struktura sieci neuronowej	161
Terminy dotyczące sieci neuronowych	162
Architektury sieci neuronowych	167
Problemy z trenowaniem sieci neuronowych	170
Modele językowe	171
Uczenie półnadzorowane	173
Uczenie nienadzorowane	173
Uczenie transferowe	174
Transformery	175
Architektura transformerów	176
Zastosowania transformerów	177
Duże modele językowe	178
Wyzwania związane z trenowaniem modeli językowych	178
Konkretne architektury modeli językowych	179
Problemy związane z używaniem GPT-3	184
Przykładowy scenariusz — projekt systemu ML/DL	
do klasyfikacji NLP w notatniku Jupytera	184
Cel biznesowy	185
Cel techniczny	185
Potok	185
Podsumowanie	188

ROZDZIAŁ 7**Duże modele językowe — teoria, projektowanie i implementacja 189**

Wymagania techniczne	190
Co to są duże modele językowe i czym różnią się od zwykłych modeli językowych?	190
Modele n-gramowe	191
Ukryte modele Markova (HMM)	191
Rekurencyjne sieci neuronowe (RNN)	191
Co wyróżnia modele LLM?	192
Powody tworzenia i używania modeli LLM	192
Lepsze wyniki	193
Szeroka generalizacja	193
Nauka na nielicznych przykładach	195
Rozumienie złożonych kontekstów	195
Wielojęzyczność	195
Generowanie tekstu podobnego do napisanego przez człowieka	196
Problemy związane z tworzeniem modeli LLM	197
Ilość danych	197
Zasoby obliczeniowe	198
Ryzyko uprzedzeń	198
Stabilność modelu	199
Interpretowalność i debugowanie	199
Wpływ na środowisko	199
Typy modeli LLM	200
Modele transformerowe	200
Przykładowe projekty nowoczesnych modeli LLM	201
GPT-3.5 i ChatGPT	201
Wstępny trening modelu językowego	204
Trening modelu nagrody	205
Dostrajanie modelu przez uczenie ze wzmacnianiem	207
GPT-4	208
LLaMA	209
PaLM	210
Narzędzia open source do RLHF	212
Podsumowanie	213
Źródła	213

ROZDZIAŁ 8**Dostęp do dużych modeli językowych —**

zaawansowana konfiguracja i integracja z RAG	215
Wymagania techniczne	216
Konfigurowanie aplikacji LLM — oparte na API modele	
o zamkniętym kodzie źródłowym	217
Wybór zdalnego dostawcy LLM	218
Inżynieria odpowiedzi i inicjalizowanie GPT	219
Eksperymentowanie z modelem GPT	220
Konfigurowanie aplikacji LLM — lokalne modele	
o otwartym kodzie źródłowym	222
Różnice między modelami o otwartym	
i zamkniętym kodzie źródłowym	222
Repozytorium modeli Hugging Face	223
Stosowanie modeli LLM z ekosystemu Hugging Face z użyciem Pythona	223
Zaawansowane projektowanie systemów — RAG i LangChain	225
Koncepcje projektowe LangChain	225
Źródła danych	225
Dane, które nie są wstępnie osadzone	227
Łańcuchy	228
Agenty	228
Pamięć długoterminowa i odwoływanie się	
do poprzednich konwersacji	229
Zapewnianie ciągłej istotności przez przyrostowe aktualizacje	
i zautomatyzowane monitorowanie	230
Omówienie prostej konfiguracji LangChain w notatniku Jupytera	231
Konfigurowanie potoku LangChain w Pythonie	231
Modele LLM w chmurze	232
AWS	233
Microsoft Azure	234
GCP	235
Podsumowanie usług chmurowych	236
Podsumowanie	237

ROZDZIAŁ 9**Eksplorowanie granic — zaawansowane zastosowania**

i innowacje w dziedzinie LLM	238
Wymagania techniczne	239
Zwiększanie skuteczności modeli LLM z użyciem RAG i LangChain —	
funkcje zaawansowane	239
Potok LangChain w Pythonie — zwiększanie skuteczności modeli LLM ...	240

Zaawansowane użycie łańcuchów	243
Zadawanie modelowi LLM pytania związanego z wiedzą ogólną	243
Nadawanie struktury danym wyjściowym — nakazywanie modelowi LLM zwrócenia wyników w określonym formacie	243
Prowadzenie płynnej konserwacji — wstawianie elementu pamięciowego w celu użycia poprzednich interakcji jako kontekstu dla następnych odpowiedzi	244
Automatyczne pobieranie informacji z różnych źródeł internetowych	247
Wyszukiwanie treści w filmach na YouTube i streszczanie ich	247
Kompresja odpowiedzi i ograniczanie kosztów użycia API	249
Kompresja odpowiedzi	250
Eksperymentowanie z kompresją odpowiedzi i ocena kompromisów ...	250
Wiele agentów — tworzenie zespołu współpracujących modeli LLM	253
Potencjalne korzyści z jednoczesnej pracy wielu agentów	254
Zespoły wielu agentów — podsumowanie	262
Podsumowanie	263

ROZDZIAŁ 10

Na fali — przeszłe, teraźniejsze i przyszłe trendy kształtowane

przez modele LLM i sztuczną inteligencję	264
Kluczowe trendy techniczne związane z modelami LLM i AI	265
Moc obliczeniowa — siła napędowa modeli LLM	265
Przyszłość mocy obliczeniowej w NLP	266
Duże zbiory danych i ich niezatarty wpływ na NLP i modele LLM	269
Cel — trening, testy porównawcze i wiedza dziedziczna	269
Wartość — niezawodność, różnorodność i efektywność	270
Wpływ — demokratyzacja, biegłość i nowe obawy	270
Ewolucja dużych modeli językowych — cel, wartość i wpływ	273
Cel — po co dążyć do większych i lepszych modeli LLM?	273
Wartość — przewaga modeli LLM	274
Wpływ — zmiana krajobrazu	274
Trendy kulturowe w NLP i modelach LLM	282
NLP i modele LLM w świecie biznesu	282
Sektory biznesu	284
Obsługa klienta — wcześnie użytkownicy	286
Zarządzanie zmianami wywołanymi przez AI	287
Trendy behawioralne wywoływane przez AI i model LLM — aspekt społeczny	291
Rosnące znaczenie asystentów osobistych	291
Łatwość komunikacji i przełamywanie barier językowych	292

Etyczne implikacje delegowanych decyzji	293
Etyka i zagrożenia — rosnące obawy związane z implementacją AI	293
Podsumowanie	296

ROZDZIAŁ 11

Okiem branży — opinie i prognozy ekspertów światowej klasy

Prezentacja ekspertów	299
Nitzan Mekel-Bobrov, PhD	299
David Sontag, PhD	299
John D. Halamka, M.D., M.S.	299
Xavier Amatriain, PhD	299
Melanie Garson, PhD	300
Nasze pytania i odpowiedzi ekspertów	300
Nitzan Mekel-Bobrov	300
David Sontag	304
John D. Halamka	308
Xavier Amatriain	313
Melanie Garson	315
Podsumowanie	319