Obraz zawierający tekst, Czcionka, logo, Grafika

Opis wygenerowany automatycznie

inż. Marta Glanowska

Nr albumu: 244815

PRACA DYPLOMOWA MAGISTERSKA

na kierunku Sztuczna Inteligencja i Uczenie Maszynowe

**Klasyfikacja typu osobowości**

**na podstawie postów internetowych**

Instytut Informatyki Stosowanej

Promotor: dr inż. Robert Susik

Łódź, 2024

**Streszczenie**

dodać

Słowa kluczowe:

**Abstract**

dodać

Keywords:

# Spis treści

[Spis treści 5](#_Toc156747550)

[Wprowadzenie 7](#_Toc156747551)

[Cel i zakres pracy 8](#_Toc156747552)

[1.1. Cel i zakres pracy 8](#_Toc156747553)

[1.2. Model osobowości MBTI 8](#_Toc156747554)

[1.3. Zbiór danych 8](#_Toc156747555)

[Przegląd literatury 10](#_Toc156747556)

[Materiały i metody 11](#_Toc156747558)

[3.1. Przygotowanie danych 11](#_Toc156747563)

[Wyniki 12](#_Toc156747564)

[Podsumowanie 13](#_Toc156747569)

[Spis tabel 14](#_Toc156747574)

[Nie można odnaleźć pozycji dla spisu ilustracji. 14](#_Toc156747575)

[Spis rysunków 15](#_Toc156747576)

[Wykaz symboli i oznaczeń 16](#_Toc156747577)

[Wykaz używanych skrótów 17](#_Toc156747578)

[Bibliografia 18](#_Toc156747583)

# Wprowadzenie

Kwestionariusz Myers-Briggs powstał w połowie ubiegłego wieku i jest jednym z najbardziej rozpowszechnionych modeli osobowości w biznesie, szczególnie w środowisku rekruterów, a także w szeroko pojętym doradztwie i coachingu. Przez lata był agresywnie reklamowany, aż na dobre osadził się w codziennych praktykach biznesowych, a także przyjął się w kulturze masowej jako sposób definicji i opisu natury poszczególnych jednostek. Mimo wątpliwości psychologów co do jego skuteczności, model ten wciąż znajduje szerokie zastosowanie we współczesnym świecie, gdyż bardzo często badani wspomnianym testem deklarują, że odnajdują się w opisie osobowości, do którego zostali zaklasyfikowali, tym samym czują się rozumiani. Tagi definiujące użytkownika z użyciem typów MBTI pojawiły się nawet w niektórych społecznościach platformy Reddit[[1]](#footnote-1), a także zostały wprowadzone do popularnych aplikacji randkowych, w celu poszukiwania jeszcze większej kompatybilności między potencjalnymi partnerami. Jedna z najbardziej popularnych stron internetowych[[2]](#footnote-2) oferujących darmowy kwestionariusz Myers-Briggs jest łatwo dostępna, atrakcyjna wizualnie, a ponadto oferuje test w wielu językach, co daje wspomnianemu modelowi znaczną przewagę nad innymi testami osobowości.

W tej pracy postawiono pytanie badawcze, czy przy użyciu modelu sztucznej inteligencji można ułatwić proces definiowania typu osobowości, oszczędzając dzięki temu czas poświęcany na wykonanie standardowego kwestionariusza. Zmiana źródła danych z odpowiedzi zaznaczanych w teście na treść postów internetowych ma dodatkową zaletę – ograniczenie lustrzanego efektu, kiedy opis danego typu zawiera informacje pochodzące z zaznaczonych odpowiedzi, przez co wydaje się być jeszcze wierniejszy. Jest to istotny aspekt, tym bardziej, że wyniki testu potrafią wahać się w zależności od okoliczności i samopoczucia danej osoby w chwili wykonywania kwestionariusza.

Ta praca ma szansę przysłużyć się do pełnego wykorzystania potencjału modelu MBTI. Skuteczność eksperymentu oznaczała zwiększenie autentyczności wyników wraz ze spotęgowaniem ich stałości, przy większej tolerancji na czynniki zewnętrzne wpływające na daną jednostkę.

Stworzony model sztucznej inteligencji może zostać wprowadzony do systemów rekrutacyjnych, by przeprowadzać wstępną ocenę osobowości kandydata na podstawie próbek tekstu, przykładowo pozyskanych z jego profesjonalnego profilu zawodowego. Takie rozpoznanie może zostać potwierdzone przez kwestionariusz na dalszym etapie rekrutacji, oszczędzając jednak czas obu stronom we wstępnym jej stadium. Należy jednak pamiętać, że kwestionariusze osobowości, chociaż mogą być pomocne w scharakteryzowaniu danego kandydata, nie powinny być czynnikiem kluczowym dla podejmowania decyzji o zatrudnieniu.

Model może być także wsparciem dla początkujących przedsiębiorców lub pracowników zastanawiających się nad swoimi naturalnymi rolami w grupie, a także rozpoznającymi swoje mocne i słabe strony. System może również zaistnieć jako wtyczka do popularnych portali społecznościowych, na których ludzie przez lata generują dziesiątki, a nawet setki postów. Takie narzędzie samorozwojowe za sprawą jednego kliknięcia byłoby w stanie określić typ osobowości danego użytkownika, a tym samym przybliżyć mu informacje na temat jego codziennych motywacji, czy sposobu komunikacji, a więc przysłużyć się do wykorzystania jego potencjału, a ponadto poprawy jego relacji z bliskimi ludźmi.

# Cel i zakres pracy

## Cel i zakres pracy

Celem pracy jest przegląd literatury, opracowanie modelu klasyfikacji, przeprowadzenie badań i analiza wyników. Zakres pracy obejmuje eksplorację możliwości zastosowania metod sztucznej inteligencji w klasyfikacji typu osobowości na podstawie postów internetowych oraz ewaluacja uzyskanych wyników.

[Na jakich efektach się skupiałam, co chciałam osiągnąć, a co nie było istotne? Na czym pracowałam (ogólnie) – zbiór danych, modele ogólnikowo.]

## Model osobowości MBTI

Powstanie testu MBTI ma swoje korzenie w teorii osobowości szwajcarskiego psychiatry i psychologa Carla Gustava Junga[[3]](#footnote-3). Jego koncepcja została rozwinięta w pierwszej połowie XX wieku przez amerykańską pisarkę i badaczkę Katharine Cook Briggs oraz jej córkę Isabel Briggs Myers, a dwadzieścia lat ich obserwacji doprowadziły ostatecznie do powstania kwestionariusza MBTI. Rzetelność i ważność modelu zostały potwierdzone i opisane w raporcie The Myers-Briggs Company[[4]](#footnote-4).

Klasyfikacja MBTI opiera się na określeniu wartości dla czterech par kategorii[[5]](#footnote-5), opisanych w *Tabeli 1*.

Tabela 1 – Opis par kategorii modelu MBTI (Źródło: opracowanie własne)

|  |  |  |
| --- | --- | --- |
|  | Para kategorii | |
| Poziom 1 – Źródło czerpania energii i komunikacja. | **Ekstrawersja (E)**  Koncentracja na świecie zewnętrznym. Komunikacja poprzez rozmowę. Towarzyskość. | **Introwersja (I)**  Koncentracja na świecie wewnętrznym. Komunikacja poprzez pismo. Powściągliwość i nieufność. |
| Poziom 2 – Sposób zbierania informacji i operowania nimi. | **Poznanie (S)**  Zorientowanie na fakty. Rzeczowość. Zapamiętywanie szczegółów. Praktyka i doświadczenie. | **Intuicja (N)**  Zorientowanie na możliwości. Kreatywność. Zapamiętywanie ogólnego obrazu. Analiza teoretyczna i inspiracja. |
| Poziom 3 – Motywacja podejmowania decyzji. | **Myślenie (T)**  Rozum ponad serce.  Rachunek „za i przeciw”.  Sprawiedliwość i logika.  Skupienie na osiągnięciu celu. | **Odczuwanie (F)**  Serce ponad rozum.  Empatia i harmonia w grupie.  Skupienie na byciu docenianym przez innych. |
| Poziom 4 – Styl życia i pracy. | **Osądzanie (J)**  Systematyczność. Planowanie. Sztywne ramy. | **Obserwacja (P)**  Spontaniczność. Swoboda.  Otwartość na modyfikacje. |

Istnieje 16 typów osobowości modelu MBTI. Czteroelementowa nazwa każdego typu jest prostym złożeniem liter z każdej z czterech par, która bardziej odpowiada naturze danej jednostki. W ten sposób, osoba o usposobieniu introwertyka (I), zorientowana na fakty i rzeczowość (S), kierująca się empatią i uczuciami (F), charakteryzująca się swobodnym stylem życia i elastycznością (P) najprawdopodobniej zostanie zaklasyfikowana do typu ISFP.

Czasami stosuje się koncentrację 16 typów w cztery grupy, aby sprawniej zorientować się w charakterystyce ich natury i pełnionych naturalnie funkcjach, są to: dążący do celu i pomysłowi Analitycy (INTJ, INTP, ENTP, ENTJ), kreatywni i zorientowani na ludzi Dyplomaci (INFJ, INFP, ENFJ, ENFP), praktyczni i skuteczni Strażnicy (ISTJ, ISFJ, ESTJ, ESFJ) oraz otwarci i spontaniczni Odkrywcy (ISTP, ISFP, ESTP, ESFP).

Wadą systemu MBTI niewątpliwie jest brak wartości pośrednich dla poszczególnych kategorii. Na przykład, w tej klasyfikacji badany nie może zostać uznany za ambiwertyka, czyli osobę posiadającą cechy z pogranicza ekstrawersji i introwersji. Może być to problematyczne w przypadku osób „uniwersalnych”, u których przychylność do jednego z biegunów w parze kategorii nie jest ewidentna. Tacy ludzie nierzadko są postrzegani jako elastyczni, potrafiący dostosować się do sytuacji, korzystający z różnych swoich atrybutów zależnie od potrzeb i okoliczności. Wykonanie wspomnianego testu może okazać się dla nich frustrującym doświadczeniem, ponieważ oscylując pośrodku dwóch biegunów kategorii, wynik klasyfikacji jest narażony na dużą zmienność wraz z kolejnymi wykonaniami testu.

## Zbiór danych

*(MBTI) Myers-Briggs Personality Type Dataset* to zbiór danych dostępny publicznie na platformie Kaggle[[6]](#footnote-6). Zawiera on 8675 wierszy, przy czym każdy z nich odpowiada ostatnim pięćdziesięciu postom udostępnionym przez danego użytkownika na forum PersonalityCafe[[7]](#footnote-7) wraz z charakteryzującym go typem osobowości w konwencji modelu MBTI. Zbiór odznacza się różnorodnością w kontekście opisywanych przez niego osób, a także samego stylu pisania. Zebranie danych z istniejącego forum, a nie w sztucznych warunkach, przyczyniło się do powstania próbek, które można określić jako autentyczne i odwzorowujące realne życie.

*Rysunek 1* przedstawia dystrybucję próbek między klasami. Każdy z wierszy zawiera maksymalnie pięćdziesiąt postów internetowych. Zbiór nie jest zbalansowany – zebrano ponad czterdziestokrotnie więcej informacji na temat osób o typie INFP niż osób scharakteryzowanych jako ESTJ czy ESFJ. Cztery najwyższe słupki wykresu opisują liczbę postów należących do osób określonych jako introwertyczne. Ponadto, około 85% danych odnosi się do użytkowników forum o typie osobowości bazującym na intuicji (składowa N), a nie zmysłach (składowa S).

Rys. 1 – Dystrybucja próbek między klasami (Źródło: opracowanie własne)



# Przegląd literatury

Klasyfikacja typów osobowości modelu MBTI na podstawie próbek tekstu przy użyciu uczenia maszynowego jest zagadnieniem chętnie podejmowanym przez badaczy na przestrzeni ostatnich lat.

W 2017 roku Rayne Hernandez i Ian Scott Knight opisali problem badawczy w swoim artykule[[8]](#footnote-8) i odnotowali, że istniejące ówcześnie rozwiązania są nieliczne. Ich praca jako jedna z pierwszych bazowała na zbiorze *(MBTI) Myers-Briggs Personality Type Dataset.* Posty internetowe zostały wstępnie przetworzone, poszczególne słowa zakodowane w przestrzeni 50-wymiarowej za pomocą algorytmu GloVe, a następnie wymodelowane za pomocą rekurencyjnych sieci neuronowych z komórkami LSTM, osobno dla każdej pary kategorii. Wspomniany artykuł dał podwaliny kolejnym pracom i okazał się jednym z najczęściej cytowanych w późniejszych latach.

We wspomnianym roku, próbę modelowania dokładnie tego samego zbioru danych podjęli również Brandon Cui i Calvin Qi[[9]](#footnote-9). Ich działania były szerzej zakrojone – zajęli się zarówno problemem klasyfikacji szesnastoklasowej jak i tworzeniem czterech odrębnych klasyfikatorów binarnych odpowiedzialnych za kolejne pary kategorii. Autorzy postanowili skupić się na drugim z wymienionych podejść, deklarując niskie wartości dokładności klasyfikatorów wieloklasowych, wynoszących poniżej 25%. W tej pracy dane tekstowe zostały wyrażone numerycznie w głównej mierze za pomocą tradycyjnej techniki worka słów (ang. *Bag of Words*).

Bianca Antonio wraz z zespołem wykonali w 2018 roku wnikliwy projekt[[10]](#footnote-10), w którym udało im się wyczerpująco opisać sam zbiór danych, przetestować kilka metod próbkowania w celu zniwelowania braku zbalansowania na przestrzeni szesnastu klas, przeprowadzić szczegółową ekstrakcję cech, włączając zliczanie wystąpień poszczególnych elementów postów, analizę sentymentu i otagowanie części mowy. Podobnie jak w poprzednich pracach, autorzy nie zdecydowali się jednak na modelowanie wieloklasowe. Po treningu modeli regresji logistycznej, maszyn wektorów nośnych (ang. *Support Vector Machines*), ekstremalnego wzmocnienia gradientu (ang. *Extreme Gradient Boosting*) oraz lasów losowych, te ostatnie osiągnęły najwyższy iloczyn dokładności dla czterech modeli binarnych, jednak wynik ten był niewiele lepszy od pracy Brandona Cui i Calvina Qi9.

Podejście zaproponowane w 2020 roku przez inną grupę badaczy[[11]](#footnote-11) zawierało porównanie skuteczności maszyn wektorów nośnych, modeli regresji logistycznej, algorytmów k najbliższych sąsiadów (ang. *K-Nearest Neighbors*) oraz lasów losowych bazując na danych zakodowanych za pomocą techniki *Word2Vec ,* oraz dodatkowych cechach wybranych po analizie wskaźników korelacji Pearsona. W tym przypadku, las losowy osiągnął 100% dokładności, jednak również był trenowany odrębnie dla każdej pary kategorii.

Interesujące wnioski wysnuli Amirhosseini i Kazemian[[12]](#footnote-12), udowadniając, że metoda ekstremalnego wzmocnienia gradientu może być skuteczniejsza niż rekurencyjna sieć neuronowa w rozwiązaniu zadanego problemu. Istotnie, cząstkowe dokładności ich modeli były wyższe niż tych w artykule z 2017 roku8.

Jedna z nowszych prac[[13]](#footnote-13) podejmująca wspomniany problem badawczy odznacza się wyższą dokładnością klasyfikacji próbek tekstowych niż wymienione powyżej. W 2022 roku zespołowi badaczy w Tajlandii udało się osiągnąć ponad 40% dokładności zarówno dla naiwnych klasyfikatorów Bayesowskich (ang. *Naive Bayes*) jak i maszyn wektorów nośnych. Przygotowanie danych do tych klasyfikatorów obejmowało zastosowanie techniki worka słów oraz ważenia częstością termów (ang. *Term Frequency-Inverse Document Frequency*). Ponadto, zaproponowana przez autorów dwukierunkowa sieć rekurencyjna z użyciem komórek LSTM osiągnęła niemal 50% skuteczności dla danych zakodowanych za pomocą tokenizera biblioteki *Keras*.

# Materiały i metody



## Przygotowanie danych

Dodać

[Konkrety – jak dzieliłam zbiór, cały pre-processing i opis metod i liczebności, jakich modeli (RNN, Word2Vec) i specyficznie architektur użyłam, szczegóły treningów, opis metryk ewaluacji i kosztu, wymagania sprzętowe i programistyczne, cytować wykorzystywane biblioteki Pythona]

# 

# Wyniki



Dodać

[Jak poradziły sobie moje architektury, co było skuteczne, a co nie (dlaczego?), ewentualne porównanie do innych architektur. Czasy treningów.]

# Podsumowanie



dodać

[Cel pracy został osiągnięty. Opisać najlepsze osiągnięte rozwiązanie. Rozważania nt. efektów w kontekście osobowości. Możliwe rozwinięcia projektu. Zastosowania w codziennym życiu.]

# 

# Spis tabel

Tabela 1 – Opis par kategorii modelu MBTI (Źródło: opracowanie własne) 9

# Spis rysunków

Rys. 1 – Dystrybucja próbek między klasami (Źródło: opracowanie własne) 12

# Wykaz symboli i oznaczeń

[O ile nie jest to inaczej wyszczególnione w tekście praca ta wykorzystuje poniższą notację:

* a (mała litera kursywą) oznacza skalar,
* **a** (mała pogrubiona litera) oznacza wektor]

# Wykaz używanych skrótów



* MBTI – Myers-Briggs Type Indicator®

# Bibliografia

1. <https://www.reddit.com/> [↑](#footnote-ref-1)
2. <https://www.16personalities.com/pl> [↑](#footnote-ref-2)
3. Carl Gustav Jung, *Typy psychologiczne,* Wydawnictwo KR 2015 [↑](#footnote-ref-3)
4. The Myers-Briggs Company, *Reliability and validity of the MBTI® instrument,* <https://eu.themyersbriggs.com/-/media/Files/PDFs/Technical-information/MBTI_reliability_and_validity_info.pdf?la=en> [↑](#footnote-ref-4)
5. Charakterystyki w *Tabeli 1* są skrótem bardziej szczegółowych opisów przedstawionych w artykule <https://potencjalosobowosci.com/mbti/>, dostęp z dnia 01.02.2024 r. [↑](#footnote-ref-5)
6. <https://www.kaggle.com/datasets/datasnaek/mbti-type>, dostęp z dnia 17.10.2023 r. [↑](#footnote-ref-6)
7. <https://www.personalitycafe.com/>, dostęp z dnia 17.10.2023 r. [↑](#footnote-ref-7)
8. Hernandez R., Knight I.S., *Predicting Myers-Briggs Type Indicator with Text Classification*, 31st Conference on Neural Information Processing Systems, 2017 [↑](#footnote-ref-8)
9. Cui B., Qi C., *Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction*, 2017 [↑](#footnote-ref-9)
10. Bianca Antonio i in., *Data Science Final Project: Myers-Briggs Prediction*, 2018, artykuł dostępny pod linkiem<https://medium.com/@bian0628/data-science-final-project-myers-briggs-prediction-ecfa203cef8>, dostęp z dnia 17.10.2023 r. [↑](#footnote-ref-10)
11. Abidin N.H.Z. i in., *Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier*, International Journal of Advanced Computer Science and Applications(IJACSA), 11(11), 2020 [↑](#footnote-ref-11)
12. Amirhosseini M. H., Kazemian H., *Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®*, *Multimodal Technologies and Interaction*, 2020 [↑](#footnote-ref-12)
13. Ontoum S., Chan J.H., *Personality Type Based on Myers-Briggs Type Indicator with Text Posting Style by using Traditional and Deep Learning*, 2022 [↑](#footnote-ref-13)