

**UNIWERSYTET GDAŃSKI**  
**WYDZIAŁ ZARZĄDZANIA**  
**Kierunek: Informatyka i ekonometria**  
**Specjalność: Analiza danych BIG DATA**

**Marcel Bilski**  
**Nr albumu 295472**

**Analiza porównawcza wybranych metod analizy emocji  
i wydźwięku w komunikatach mediów  
społecznościowych**

**Praca magisterska napisana  
w Katedrze Statystyki Wydziału Zarządzania  
pod kierunkiem dr. hab. Krzysztofa Najmana, prof. UG**

**Gdańsk 2025**

## Spis treści

<b>WSTĘP.....</b>	3
<b>ROZDZIAŁ 1 .....</b>	4
1.1.    Rola emocji w życiu człowieka .....	4
1.2.    Klasyfikacja emocji według Plutchika .....	6
1.2.1.    Postulaty Teorii Emocji Według Plutchika .....	6
1.2.2.    Osiem emocji podstawowych według Plutchik .....	8
1.2.3.    Koło emocji według Plutchik .....	9
1.2.4.    Emocje wtórne .....	10
1.3.    Rola emocji w społeczeństwie, biznesie i organizacji – wpływanie na emocje pracowników, przedsiębiorstw i klientów .....	10
1.3.1.    Emocje w miejscu pracy: od zaniedbania do uznania .....	11
1.4.    Humanizacja biznesu poprzez emocje .....	11
1.4.1.    Emocje a etyka biznesowa .....	12
1.4.2.    Wpływ emocji na klientów .....	12
1.4.3.    Zarządzanie emocjami w organizacji .....	13
<b>ROZDZIAŁ 2 .....</b>	14
2.1.    Text mining .....	14
2.2.    Analiza sentymentów .....	16
2.3.    Metody w analizie sentymentów .....	17
2.3.1.    Metody słownikowe .....	18
2.3.2.    Uczenie maszynowe .....	19
<b>ROZDZIAŁ 3 .....</b>	21
3.1. Opis badań i danych .....	21
3.2. Badania .....	22
3.2.1. Podział sentymentów .....	23
3.2.2. Frekwencja emocji i sentymentów .....	24
3.2.3. Macierz sentymentów w NRC vs polarność uczeniu maszynowym .....	26
3.2.4. Rozkład emocji i sentymentów w zbiorze danych .....	27
3.2.5. Macierz współwystępowania .....	29
<b>ZAKOŃCZENIE .....</b>	31
<b>Bibliografia .....</b>	32
<b>SPIS TABEL .....</b>	35
<b>SPIS RYSUNKÓW .....</b>	35
<b>SPIS ZAŁĄCZNIKÓW .....</b>	35

## **WSTĘP**

Współczesny świat, w którym komunikacja coraz częściej przenosi się do przestrzeni cyfrowej, stawia przed badaczami nowe wyzwania związane z rozumieniem emocji wyrażanych w przekazach medialnych. W dobie szybkiego obiegu informacji i zautomatyzowanego przetwarzania danych rośnie znaczenie analizy emocji i sentymenu, które pozwalają w sposób ilościowy i jakościowy badać emocjonalny ton wypowiedzi.

Celem niniejszej pracy jest analiza porównawcza wybranych metod analizy emocji i wydźwięku komunikatów w mediach społecznościowych. Badania opierają się na popularnym zbiorze danych Sentiment140, który został zbudowany z myślą o automatycznej klasyfikacji emocjonalnego wydźwięku wypowiedzi publikowanych na platformie X (dawniej Twitter). Dane te zostały oznaczone metodą *distant supervision* – poprzez obecność określonych emotikonów w treści, umożliwiając przypisanie sentymenu bez konieczności ręcznej klasyfikacji.

W części teoretycznej pracy przedstawiono psychologiczne podstawy analizy emocji, koncentrując się na teorii emocji Plutchika, która służy jako punkt wyjścia do interpretacji danych. Omówiono także rolę emocji w życiu społecznym, biznesie i komunikacji – ze szczególnym uwzględnieniem ich znaczenia w miejscu pracy, kontaktach z klientami oraz strategiach komunikacyjnych organizacji.

W części badawczej dokonano analizy emocji i sentymenu za pomocą NRC Emotion Lexicon (EmoLex) – słownikowego narzędzia umożliwiającego przypisanie konkretnych emocji i ogólnego wydźwięku do słów występujących w tekście. Analiza została przeprowadzona na próbce 2000 tweetów ze zbioru Sentiment140. Pozwoliło to zbadać, które emocje dominują w wypowiedziach użytkowników, jak często występują wspólnie, a także jak rozkładają się pomiędzy poszczególnymi autorami. Celem badania jest ocena spójności wyników uzyskanych obiema metodami i identyfikacja różnic w rozpoznawaniu emocji i wydźwięku.

Praca ma charakter interdyscyplinarny – łączy zagadnienia z zakresu psychologii emocji, językoznawstwa komputerowego oraz analityki danych. Jej celem jest nie tylko porównanie skuteczności różnych metod analizy emocji, ale także refleksja nad tym, w jakich obszarach – takich jak badania opinii publicznej, zachowania konsumenckie czy komunikacja marek – metody te mogą znaleźć praktyczne zastosowanie.

# **ROZDZIAŁ 1**

## **1.1. Rola emocji w życiu człowieka**

Emocje odgrywają istotną rolę w naszym życiu człowieka. Wpływają na nasze zachowania, podejmowane przez nas decyzje, nasze relacje społeczne oraz zdrowie psychiczne. Natura emocji i ich rola była i jest ciągle rozważana przez naukowców, filozofów i artystów. Można ją rozważyć pod względami społecznymi, kulturowymi, biologicznymi czy językowymi. Postrzeganie roli emocji jest uzależnione od warunków zewnętrznych, kształtujących spojrzenie na nią przez ludzi ze względu na wykształcenie osoby, jej wiek i status społeczny oraz epokę w jakiej ta osoba żyje.

Czym więc dokładnie są emocje i jaką rolę odgrywają w życiu człowieka? Istnieje wiele definicji w literaturze fachowej pojęcia **emocje**, więc nie jednej jedynej ogólnej definicji tego pojęcia. Samo słowo emocja pochodzi od łacińskiego słowa *e mouere*, co oznacza „w ruchu”. *Encyklopedia PWN* definiuje emocje jako: „*system obejmujący uczucia, uznawane za składnik subiektywny, pobudzenie fizjologiczne wraz ze specyficzną ekspresją oraz zmiany zachowania*”.<sup>1</sup> Zaś Dariusz Doliński twierdzi, że emocje można rozumieć jako: „*subiektywny stan psychiczny, uruchamiający priorytet dla związanego z nim programu działania. Jej odczuwaniu towarzyszą zwykle zmiany somatyczne, ekspresje mimiczne i pantomimiczne oraz zachowania*”. Można w takim przypadku zrozumieć emocje jako ‘tryb’ w programie komputerowym, którego ‘włączenie’ lub zmiana z innego trybu (innych emocji) w wyniku doświadczenia zewnętrznych lub wewnętrznych czynników, skutkuje w zmianie programu działania, czemu towarzyszą zmiany cielesne, zmiana mowy ciała i zmiany komunikacji niewerbalnej.<sup>2</sup>

W filozofii emocje były przez dług okres postrzegany jako najmniej zrozumiane i nieuchwytny składnik do całościowej klasyfikacji i zrozumienia, co wynikało z ich złożoności – budowy emocji, źródeł ich powstania oraz relacji jakie mają z innymi elementami ludzkiej natury. Istniały tylko przypuszczenia, że dla emocji istnieje relacja pomiędzy aspektami mentalnymi a fizjologicznymi człowieka. Filozofia nowożytna zbudowana na gruncie

---

<sup>1</sup> Emocje, <https://encyklopedia.pwn.pl/szukaj/emocje> [dostęp: 05.02.2025]

<sup>2</sup> Doliński D., *Psychologia. Podręcznik akademicki*, t. 2, Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2004, s. 322.

Kartezańskiego podejścia pozwoliła spojrzeć na emocje i ich istotę w inny sposób. Chociaż za ojca fizjologicznej teorii emocji uznaje się W. Jamesa z jego Teorią Jamesa-Langego, to w niektórych częściach teoria W. Jamesa przypomina ujęcia Kartejusza, o czym też mówi G. Hatfield<sup>3</sup>. Chociaż teoria W. James jest w pełni fizjologiczna, gdzie Kartejusz pisze o emocjach jako powiązaniu duszy i ciała, a według W. James emocje są zdefiniowane jako doświadczenia psychiczne są tylko etykietą jaką nadajemy naszym reakcjom fizjologicznym. Można więc uznać Kartezańską koncepcję emocji jako co najwyżej prekursora teorii W. James.<sup>4</sup>

Joanna Krzemkowska-Saja w swoim artykule, uważa Kartezańską koncepcję jako kompromis między stanowiskami fizjologicznych i poznawczy teoriami emocji<sup>5</sup>. Podejście do badania emocji Kartejusza opiera się na jego teorii spostrzegania wewnętrznego, która jest równoznaczna z dzisiejszą teorią introspekcji. W ramach spostrzegania wewnętrznego zakłada się, że podmiot ma uprzywilejowany dostęp do własnych stanów mentalnych, co dostarcza nam bezpośrednich danych na temat stanów mentalnych. Te dane są niepowątpliwne, co oznacza jeśli czujemy strach, to nie mogę się mylić na temat, że czuję ten strach. Według ujęcia Kartejusza dane pozyskane w spostrzegania wewnętrznego stanowią podstawę dla wiedzy o świecie nas otaczającym jak również wiedzę na temat naszego życia psychicznego, duchowego i emocjonalnego<sup>6</sup>. Jednak sam Kartejusz uznał, że introspekcja nie jest w stanie dać nam w pełni dostępu do naszych stanów emocjonalnych, czego przyczyną są bodźce mechaniczne i fizjologiczne, do których umysł nie ma dostępu. Według Kartezańskiej metody, emocje mają podłożę stanowiące całkowicie mechaniczne procesy, które zachodzą w ciele. Emocje w takim wypadku można uznać za najbardziej „ucielesnione” stany mentalne. Emocje jako stany o podłożu fizjologicznym są mechanizmami mającymi ułatwić ciału zdobycie rzeczy koniecznych i uniknięcie grożących mu niebezpieczeństw.

---

<sup>3</sup> Hatfield G., *Did Descartes Have a Jamesian Theory of the Emotions?*, “Philosophical Psychology”, 4 (2007), Vol. 20, s. 413-440

<sup>4</sup> Kalat, James W., *Biologiczne podstawy psychologii*, Wydawnictwo naukowe PWN, 2006, str. 358-361

<sup>5</sup> Krzemkowska-Saja J., Namiętności duszy a współczesny spór o naturę emocji. „Filo–Sofija”, 2012, Nr 17(2012/2), s. 137-145

<sup>6</sup> Judycki S., *Introspekcja jako problem filozoficzny*, „Roczniki Filozoficzne”, 50 (2002), z. 1, s. 5-6

## **1.2. Klasyfikacja emocji według Plutchika**

### **1.2.1. Postulaty Teorii Emocji Według Plutchika**

Robert Plutchik w swojej psycho ewolucyjnej teorii emocji przedstawił dziesięć postulatów, które stanowią podstawę jego modelu emocji. Te postulaty mają na celu wyjaśnienie natury emocji, ich funkcji adaptacyjnych oraz relacji między emocjami a innymi dziedzinami psychologii, takimi jak osobowość, diagnoza kliniczna i mechanizmy obronne ego.<sup>7</sup> Oto główne postulaty teorii Plutchika:

#### **Postulat 1: Uniwersalność emocji**

Koncepcja emocji ma zastosowanie na wszystkich poziomach ewolucyjnych i dotyczy zarówno zwierząt, jak i ludzi. Plutchik podkreśla, że emocje są uniwersalne i można je zaobserwować u różnych gatunków, co wskazuje na ich ewolucyjne korzenie.

#### **Postulat 2: Ewolucyjna historia emocji**

Emocje mają historię ewolucyjną i przybierają różne formy ekspresji u różnych gatunków. Plutchik twierdzi, że emocje ewoluowały jako mechanizmy adaptacyjne, które pomagają organizmom radzić sobie z wyzwaniami środowiskowymi.

#### **Postulat 3: Funkcja adaptacyjna emocji**

Emocje pełnią rolę adaptacyjną, pomagając organizmom radzić sobie z kluczowymi problemami przetrwania, takimi jak zagrożenia, poszukiwanie pożywienia czy reprodukcja. Plutchik uważa, że emocje są reakcjami na sytuacje, które mają znaczenie dla przetrwania.

#### **Postulat 4: Wspólne elementy emocji**

Pomimo różnic w ekspresji emocji u różnych gatunków, istnieją pewne wspólne elementy lub wzorce prototypowe, które można zidentyfikować. Plutchik twierdzi, że emocje mają uniwersalne cechy, które są obecne u różnych organizmów.

#### **Postulat 5: Podstawowe emocje**

---

<sup>7</sup> Plutchik, R., *Emotion: A psychoevolutionary synthesis*. "The American Journal of Psychology". 1980, nr 93(4) s. 8-9

Istnieje niewielka liczba podstawowych, pierwotnych emocji, które są uniwersalne i stanowią podstawę dla bardziej złożonych stanów emocjonalnych. Plutchik wyróżnia osiem podstawowych emocji: radość, smutek, złość, strach, zaskoczenie, oczekiwanie, akceptacja i wstręt.

#### **Postulat 6:** Emocje pochodne/wtórne

Wszystkie inne emocje są mieszanymi lub pochodnymi stanami, które powstają w wyniku kombinacji, mieszanin lub związków emocji podstawowych. Plutchik podkreśla, że emocje wtórne, takie jak miłość czy zazdrość, są wynikiem połączenia dwóch lub więcej emocji podstawowych.

#### **Postulat 7:** Hipotetyczny charakter emocji

Podstawowe emocje są hipotetycznymi konstruktami lub zidealizowanymi stanami, których właściwości i charakterystyki można wywnioskować na podstawie różnych rodzajów dowodów. Plutchik uważa, że emocje nie są bezpośrednio obserwowlane, ale można je wywnioskować na podstawie zachowań, reakcji fizjologicznych i verbalnych opisów.

#### **Postulat 8:** Biegunowość emocji

Podstawowe emocje można konceptualizować jako pary przeciwnieństw. Na przykład radość jest przeciwnieństwem smutku, a złość przeciwnieństwem strachu. Plutchik podkreśla, że emocje mają charakter bipolarny, co oznacza, że każda emocija ma swoje przeciwnieństwo.

#### **Postulat 9:** Podobieństwo emocji

Wszystkie emocje różnią się stopniem podobieństwa do siebie. Plutchik twierdzi, że emocje, które są podobne, znajdują się blisko siebie na kole emocji, podczas gdy emocje przeciwwstawne znajdują się naprzeciwko siebie.

#### **Postulat 10:** Intensywność emocji

Każda emocija może istnieć w różnym stopniu intensywności lub poziomu pobudzenia. Plutchik podkreśla, że emocje mogą być bardziej lub mniej intensywne, co wpływa na ich ekspresję i funkcję adaptacyjną.

Te postulaty stanowią podstawę teorii Plutchika i pozwalają na zrozumienie, jak emocje funkcjonują w kontekście ewolucyjnym, jak są ze sobą powiązane oraz jak wpływają na inne aspekty psychologii, takie jak osobowość czy mechanizmy obronne. Plutchik podkreśla, że

emocje są nie tylko subiektywnymi doświadczeniami, ale również mają kluczowe znaczenie dla przetrwania i adaptacji organizmów.

### 1.2.2. Osiem emocji podstawowych według Plutchik

Robert Plutchik w swojej teorii psycho ewolucyjnej wyróżnił osiem podstawowych emocji, które uważa za uniwersalne i ewolucyjnie uwarunkowane. Te emocje są związane z adaptacyjnymi funkcjami, które pomagają organizmom przetrwać i radzić sobie z wyzwaniami środowiskowymi.<sup>8</sup> Osiem podstawowych emocji to:

1. **Radość** – związana z odczuwaniem przyjemności i satysfakcji. Jej funkcją adaptacyjną jest wzmacnianie więzi społecznych i zachęcanie do powtarzania korzystnych zachowań.
2. **Smutek** – związany z utratą lub brakiem. Jego funkcją jest sygnalizowanie potrzeby reintegracji lub odzyskania utraconego obiektu.
3. **Gniew** – związana z frustracją lub zagrożeniem. Jej funkcją jest mobilizacja organizmu do obrony lub ataku.
4. **Strach** – związany z zagrożeniem lub niebezpieczeństwem. Jego funkcją jest przygotowanie organizmu do ucieczki lub uniknięcia zagrożenia.
5. **Zaskoczenie** – związane z nagłymi, nieoczekiwanyimi zdarzeniami. Jego funkcją jest przygotowanie organizmu do szybkiej reakcji na nowe sytuacje.
6. **Przeczuwanie** – związane z antycypacją przyszłych zdarzeń. Jego funkcją jest przygotowanie organizmu do przyszłych wyzwań lub możliwości.
7. **Zaufanie** – związana z odczuwaniem bliskości i przynależności. Jej funkcją jest wzmacnianie więzi społecznych i współpracy.
8. **Wstręt** – związany z odrzuceniem czegoś szkodliwego lub nieprzyjemnego. Jego funkcją jest ochrona organizmu przed szkodliwymi substancjami lub sytuacjami.

---

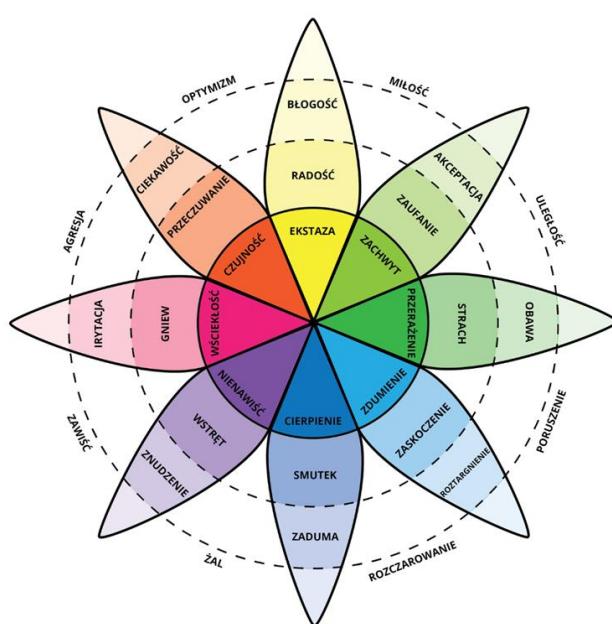
<sup>8</sup> Plutchik R., *The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice*, "American Scientist", 2001, nr 89(4), s. 349

Te emocje są uważane za podstawowe, ponieważ są uniwersalne, mają wyraźne funkcje adaptacyjne i są obecne u różnych gatunków zwierząt, co wskazuje na ich ewolucyjne korzenie.

### **1.2.3. Koło emocji według Plutchik**

Plutchik zaproponował model koła emocji (tzw. koło Plutchika), który ilustruje relacje między podstawowymi emocjami oraz ich intensywnością. Model ten jest analogiczny do koła kolorów, gdzie emocje są ułożone wokół okręgu, a ich intensywność jest reprezentowana przez odległość od środka koła.

- **Podstawowe emocje** znajdują się na zewnętrznym obwodzie koła, a ich intensywność maleje w miarę zbliżania się do środka. Na przykład, intensywna złość może przejść w irytację, a następnie w łagodne niezadowolenie.
  - **Emocje przeciwwstne** znajdują się naprzeciwko siebie na kole. Na przykład, radość jest przeciwwstwem smutku, a złość przeciwwstwem strachu.
  - **Emocje podobne** znajdują się blisko siebie na kole, co oznacza, że są ze sobą powiązane. Na przykład, strach i zaskoczenie są blisko siebie, ponieważ obie są reakcjami na nagłe zmiany w środowisku.



### Rysunek 1 Koło emocji Roberta Plutchika

Źródło: Englishsquare.pl sp. z o.o., licencja: CC BY-SA 3.0.

Model koła emocji pozwala na wizualizację relacji między emocjami oraz ich gradacji intensywności, co ułatwia zrozumienie, jak różne emocje mogą się ze sobą łączyć i przekształcać.

#### **1.2.4. Emocje wtórne**

Plutchik wprowadził również koncepcję emocji wtórnych, które powstają w wyniku mieszania się podstawowych emocji. Emocje wtórne są bardziej złożone i często występują w codziennym życiu. Przykłady emocji wtórnych to:

- **Miłość** – mieszanka radości i zaufania. Miłość jest związana z odczuwaniem bliskości i przywiązania do innych.
- **Zazdrość** – mieszanka gniewu i strachu. Zazdrość pojawia się, gdy obawiamy się utraty czegoś cennego, co może prowadzić do frustracji i gniewu.
- **Wstyd** – mieszanka strachu i smutku. Wstyd jest związany z poczuciem winy lub nieadekwatności.
- **Duma** – mieszanka radości i akceptacji. Duma pojawia się, gdy odczuwamy satysfakcję z własnych osiągnięć lub uznania ze strony innych.

Emocje wtórne są wynikiem połączenia dwóch lub więcej podstawowych emocji, co prowadzi do powstania bardziej złożonych stanów emocjonalnych. Plutchik uważa, że te emocje są równie ważne w życiu codziennym, jak emocje podstawowe, ponieważ odzwierciedlają złożoność ludzkich doświadczeń.

### **1.3. Rola emocji w społeczeństwie, biznesie i organizacji – wpływanie na emocje pracowników, przedsiębiorstw i klientów**

Emocje odgrywają kluczową rolę w funkcjonowaniu biznesu i organizacji, wpływając na relacje między pracownikami, zarządzanie, a także na postrzeganie marki przez klientów. W ostatnich latach coraz więcej uwagi poświęca się emocjom w kontekście zarządzania, etyki biznesowej oraz wpływu na efektywność organizacyjną. W niniejszym podrozdziale omówimy,

jak emocje wpływają na pracowników, przedsiębiorstwa i klientów, oraz jak można je wykorzystać do budowania bardziej humanitarnych i efektywnych organizacji.

### **1.3.1. Emocje w miejscu pracy: od zaniedbania do uznania**

Przez wiele lat emocje w miejscu pracy były tematem marginalizowanym w badaniach nad zachowaniami organizacyjnymi. Jak zauważa Paul M. Muchinsky<sup>9</sup>, emocje były postrzegane jako niepożądane czynniki, które mogą zakłócać racjonalne decyzje i procesy zarządcze. W latach 50. XX wieku dominował pogląd, że skuteczny menedżer powinien być przede wszystkim logiczny i racjonalny, a emocje traktowano jako oznakę słabości lub niestabilności. Jednak wraz z rozwojem psychologii emocji i badań nad zarządzaniem, zaczęto dostrzegać, że emocje są nieodłącznym elementem ludzkiego doświadczenia, także w kontekście zawodowym.

Emocje w miejscu pracy mogą być zarówno pozytywne, jak i negatywne. Pozytywne emocje, takie jak radość, duma czy satysfakcja, mogą zwiększać zaangażowanie pracowników, poprawiać ich wydajność i sprzyjać budowaniu silniejszych relacji w zespole. Z kolei negatywne emocje, takie jak gniew, frustracja czy lęk, mogą prowadzić do spadku motywacji, konfliktów, a nawet wypalenia zawodowego. Dlatego zrozumienie i zarządzanie emocjami w organizacji stało się kluczowym elementem współczesnego zarządzania.

## **1.4. Humanizacja biznesu poprzez emocje**

Yotam Lurie w swoim artykule „Humanizing Business through Emotions” podkreśla, że emocje są nieodzownym elementem humanizacji biznesu. Zarządzanie oparte na emocjach może prowadzić do tworzenia bardziej empatycznych i etycznych organizacji, w których pracownicy czują się doceniani i wspierani. Lurie wskazuje, że emocje, takie jak współczucie, troska czy empatia, są kluczowe dla budowania relacji opartych na zaufaniu i wzajemnym

---

<sup>9</sup> Muchinsky P., *Emotions in the Workplace: The Neglect of Organizational Behavior*, “Journal of Organizational Behavior”, 2000, nr 21(7), s. 801-805

szacunku. W kontekście biznesowym, emocje te mogą przełożyć się na lepsze zarządzanie zespołem, zwiększenie lojalności pracowników oraz poprawę atmosfery w miejscu pracy.<sup>10</sup>

Przykładem może być kontrast między dwoma typami menedżerów: Larrym, który jest bezwzględny i skoncentrowany wyłącznie na zyskach, a Jorgym, który troszczy się o swoich pracowników i społeczność, w której działa firma. Lurie argumentuje, że menedżerowie, którzy potrafią okazywać emocje i angażować się w relacje z pracownikami, są bardziej skutecznymi w budowaniu długotrwałych wartości organizacyjnych.

#### **1.4.1. Emocje a etyka biznesowa**

Robert C. Solomon w swojej pracy „The Moral Psychology of Business”<sup>11</sup> podkreśla, że emocje są nieodłącznym elementem etyki biznesowej. Współczucie, troska i empatia są nie tylko cnotami moralnymi, ale także praktycznymi narzędziami, które mogą poprawić funkcjonowanie organizacji. Solomon argumentuje, że emocje, takie jak współczucie, są kluczowe dla budowania sprawiedliwości w organizacji. Sprawiedliwość nie polega wyłącznie na stosowaniu abstrakcyjnych zasad, ale na zrozumieniu i uwzględnieniu emocjonalnych potrzeb pracowników i innych interesariuszy.

Solomon wskazuje również, że emocje mogą być motorem działań etycznych. Na przykład, współczucie może skłonić menedżera do podjęcia decyzji, które będą korzystne dla pracowników, nawet jeśli w krótkim okresie nie przyniosą one korzyści finansowych. W ten sposób emocje mogą stać się podstawą do budowania bardziej humanitarnych i etycznych organizacji.

#### **1.4.2. Wpływ emocji na klientów**

Emocje odgrywają również kluczową rolę w relacjach z klientami. Badania nad zachowaniami konsumentów pokazują, że emocje wywołane przez reklamy, opakowania czy

---

<sup>10</sup> Lurie Y., *Humanizing Business through Emotions: On the Role of Emotions in Ethics*, “Journal of Business Ethics”, 2004, nr 49(1), s. 1

<sup>11</sup> Solomon R., *The Moral Psychology of Business: Care and Compassion in the Corporation*, “Business Ethics Quarterly”, 1998, nr 8(3), s. 516

obsługę klienta mają znaczący wpływ na decyzje zakupowe. Morris B. Holbrook i Rajeev Batra w swoich badaniach nad rolą emocji w reklamie wykazali, że emocje są mediatorami między treścią reklamy a postawami wobec marki. Pozytywne emocje, takie jak radość czy nostalgia, mogą zwiększać atrakcyjność marki, podczas gdy negatywne emocje, takie jak strach czy gniew, mogą prowadzić do unikania produktu lub usługi.<sup>12</sup>

Dlatego firmy coraz częściej wykorzystują emocje w strategiach marketingowych, tworząc kampanie, które wywołują silne reakcje emocjonalne. Przykładem może być wykorzystanie humoru w reklamach, które budują pozytywne skojarzenia z marką, lub kampanie społeczne, które wywołują współczucie i skłaniają do działania.

#### **1.4.3. Zarządzanie emocjami w organizacji**

Zarządzanie emocjami w organizacji wymaga zarówno zrozumienia ich roli, jak i wdrożenia odpowiednich strategii. Paul M. Muchinsky sugeruje, że organizacje powinny zwracać większą uwagę na emocje pracowników, zarówno te pozytywne, jak i negatywne. Wprowadzenie programów wsparcia emocjonalnego, szkoleń z zarządzania stresem czy budowanie kultury organizacyjnej opartej na empatii może znaczco poprawić atmosferę w miejscu pracy i zwiększyć efektywność zespołów.<sup>13</sup>

Jednym z kluczowych pojęć w zarządzaniu emocjami jest inteligencja emocjonalna, która obejmuje umiejętność rozpoznawania, rozumienia i zarządzania własnymi emocjami oraz emocjami innych osób. Daniel Goleman, autor koncepcji inteligencji emocjonalnej, wskazuje, że menedżerowie o wysokim poziomie inteligencji emocjonalnej są bardziej skuteczni w zarządzaniu zespołami, rozwiązywaniu konfliktów i budowaniu zaangażowania pracowników.<sup>14</sup>

---

<sup>12</sup> Holbrook M., Batra R., *Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising*, "Journal of Consumer Research", 1987, nr 14(3), s. 404

<sup>13</sup> Muchinsky P., *Emotions in the Workplace...*, dz. cyt., s. 802

<sup>14</sup> Tamże, s. 804

## **ROZDZIAŁ 2**

Ten rozdział poświęcony jest wprowadzeniu w obszerną dziedzinę eksploracji danych tekstowych (Text Mining) oraz szczegółowemu omówieniu analizy sentymentów jako jej kluczowej metody. Przedstawione zostaną podstawowe koncepcje, historia rozwoju, typowe zadania i zastosowania Text Miningu, a także szczegółowe metody analizy sentymentów, w tym podejścia słownikowe i bazujące na uczeniu maszynowym.

### **2.1. Text mining**

Pojęcie text mining zaczęło zyskiwać na popularności pod koniec lat dziewięćdziesiątych XX wieku. Jednym z proponentów tego pojęcia jest Marti A. Hearst, która zdefiniowała text mining jako badanie eksploracyjne w taki o to sposób: „*Eksploracja danych tekstowych jest procesem eksploracyjnej analizy danych, który prowadzi do odkrycia nieznanych dotąd informacji lub odpowiedzi na pytania, na które odpowiedź nie jest obecnie znana*”.<sup>15</sup>

Text mining ma charakter interdyscyplinarny. Wynika to z tego, że korzysta informacjami pozyskanyymi z data miningu, przetwarzania języka naturalnego, uczenia maszynowego, statystyki, lingwistyki oraz informatyki. Chociaż jest wiele podobieństw pomiędzy text miningiem a podanymi dziedzinami, jest też wiele różnic.<sup>16</sup>

Text mining i data mining mają na celu pomóc analitykowi odkrycie nieznanych prawidłowości, używając do tego eksploracji dużych zasobów danych. Jednak podejście tych dziedzin do przetwarzania tych danych jest inne. Narzędzia zbudowane na data miningu, informacje zostają uzyskane z ustrukturyzowanych baz danych. Natomiast w procesie text miningu, dane są uzyskane z nieustrukturyzowanych dokumentów tekstowych, które są zapisane w języku naturalnym, używanym do komunikacji interpersonalnej.<sup>17</sup>

---

<sup>15</sup> Hearst M., *Untangling Text Data Mining*, School of Information Management & Systems University of California, Berkeley, 1999, s. 5

<sup>16</sup> Lula P., *Text mining jako narzędzie pozyskiwania informacji z dokumentów tekstowych*, Statsoft., Kraków, 2005, s. 68

<sup>17</sup> Kuligowska K., Lasek M., Eksploracja danych tekstowych (text mining) w przedsiębiorstwie, „Polskie Stowarzyszenie Zarządzania Wiedzą”, 2008, nr 13, s. 2

Text mining również korzysta z rozwiązań zaczerpniętych z przetwarzania języka naturalnego (NLP). Przetwarzanie języka naturalnego to próba wyodrębnienia pełniejszej reprezentacji znaczenia z tekstu. Można to ująć jako ustalenie odpowiedzi na następujące pytania: kto, co, kiedy, gdzie, jak i dlaczego zrobił. NLP zazwyczaj wykorzystuje koncepcje językowe, takie jak część mowy i strukturę. W tym celu NLP używa leksykonu słów, ich znaczenia, funkcji gramatycznych oraz zasad gramatyki. Text mining używa narzędzi zaczerpniętych z NLP, przykładowo: tokenizacji, lematyzacji, analizy składniowej i semantycznej, w celu przekształcenia tekstu na dane.<sup>18</sup>

Termin uczenie maszynowe odnosi się do automatycznego wykrywania znaczących wzorców w danych poprzez analizę, po czym wspomniane wzorce są ujęte w postaci modelu, który system komputerowy może w łatwy sposób wykorzystać<sup>19</sup>. Dziedzina uczenia maszynowego zajmuje się pytaniem, jak skonstruować programy komputerowe, które automatycznie poprawiają się wraz z doświadczeniem<sup>20</sup>. W ciągu ostatnich kilku dekad stało się ono powszechnym narzędziem w niemal każdym zadaniu, które wymaga ekstrakcji informacji z dużych zbiorów danych<sup>21</sup>. Przykładami modeli najczęściej stosowanych są sieci neuronowe, algorytmy genetyczne oraz drzewa decyzyjne. Uczenie maszynowe jest często używane w text miningu, chociaż nie zawsze. W innych wypadkach stosowane jest podejście skupiające się na próbie ekstrakcji informacji za pomocą metod statystycznych klasycznych.

Text mining jest stosunkowo młodą dziedziną wielodyscyplinarną. Jak każda inna dziedzina używa również osiągnięć statystyki, która pozwala opisywać, wyjaśniać i prawidłowości i prognozy<sup>22</sup>. Często używanymi metodami z statystyki są metody statystyki opisowej, łańcuchy Markowa oraz modele probalistyczne. Modelowanie statystyczne w text miningu pozwala opisywać zawartość tekstu w sposób przydatny do umożliwienia dalszego przetwarzanie.<sup>23</sup>

Pod wieloma aspektami text mining jest związany z informatyką. Komputery mają długą historię związaną z przetwarzaniem danych tekstowych. Algorytmy pozwalające na przetwarzanie danych były tematem rozważań osób zajmujących się algorytmiką. Obecnie text

---

<sup>18</sup> Kao A., Poteet S., *Natural Language Processing and Text Mining*, Springer, United States of America, 2007, s. 1

<sup>19</sup> Lula P., *Text mining jako narzędzie...*, dz. cyt., s. 68

<sup>20</sup> Mitchell T., *Machine Learning*, McGraw-Hill Science/Engineering/Math, United States of America, 1997, s. 2

<sup>21</sup> Shalev-Shwartz S., Ben-David S., *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, United States of America, 2014, s. 20

<sup>22</sup> Lula P., *Text mining jako narzędzie...*, dz. cyt., s. 71

<sup>23</sup> Gładysz A., *Przegląd zastosowań analizy text miningowej*, „Autobusy”, 2016, nr 12, s. 1743

mining używa osiągnięcia informatyki związane z projektowaniem, analizą i przetwarzaniem struktur danych. Osiągnięcia zapożyczone z informatyki pozwalają rozwiązać problemy wynikające z przechowywania informacji pozyskanych z danych badawczych. Na granicy tych dwóch dziedzin istnieją takie same problemy skupiające się na przechowywaniu dużej ilości danych oraz operacji indeksowania danych.

## 2.2. Analiza sentymentów

Opinie mają duże znaczenie dla wszystkich działań wykonywanych przez ludzi i mają wpływ na ich zachowania. Uwarunkowują one nasze postrzeganie rzeczywistości i ludzi wokół nas. Gdy występuje sytuacja, gdzie musimy podjąć decyzje, wtedy zwracamy się do innych osób po ich opinie. Opinie mają nie tylko wpływ na indywidualnych ludzi, ale na całe grupy. Opinie i powiązane z nimi pojęcia, takie jak sentymenty, postawy i emocje są przedmiotem badań analizy sentymentów. Nagły rozwój tej dziedziny wystąpił w tej samej chwili, co rozwój mediów społecznościowych. Dane zebrane z serwisów Internetów pozwalają zilustrowanie opinii milionów osób w formie komentarzy i postów.

Omawiana dziedzina, najczęściej jest nazywana analiza sentymentu (Sentiment analysis), jest używane zamiennie z sformułowaniem analiza opinii (Opinion mining)<sup>24</sup>. W obu przypadkach pierwszy człon tych sformułowań ma związek z automatycznymi i półautomatycznymi metodami analizy treści<sup>25</sup>. Zadaniem tych dwóch metod jest identyfikowanie i klasyfikowanie badanych wypowiedzi na podstawie użytych w tekście słów i fraz. W przypadku analizy sentymentu, jest wyraźne skupienie na słowa nacechowane emocjonalnie, a analizę opinii można uznać za rozszerzenie analizy sentymentu, ponieważ analiza sentymentu skupia się nie tylko na opinii zidentyfikowanych na podstawie tematu czy formy wypowiedzi, ale również ze względu na obiekt, którego opinia dotyczy, przykładowo styl sformułowanej wypowiedzi<sup>26</sup>. Analiza sentymentu jest więc jednym z elementów obszaru badanego przez analizę opinii. Podstawowym zadaniem analizy opinii jest klasyfikacja

---

<sup>24</sup> Liu B., *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, United States of America, 2012, s. 7

<sup>25</sup> Tomanek K., *Metodyka dla analizy treści w projektach stosujących techniki text mining i rozwiązania CAQDAS piątej generacji*, „Przegląd Socjologii Jakościowej”, 2017, nr 13(2), s. 129

<sup>26</sup> Tamże, s. 129

polarności<sup>27</sup>. Wynika to z tego, że w podstawowym założeniu analizy sentymantu, wypowiedzi są klasyfikowane w trzy różne kategorie: pozytywne, negatywne i neutralne<sup>28</sup>. Analiza sentymantu ma za zadanie ustalić emocjonalną postawę wobec danego obiektu czy zdarzenia.

### **2.3. Metody w analizie sentymantu**

Analiza sentymantu, inaczej nazywana analizą wydźwięku, wykorzystuje przetwarzanie języka naturalnego (z języka angielskiego, Natural Language Processing). W wyniku dojrzewania tej dziedziny, zrodziło się wiele technik, które różnią się od siebie zawansowaniem technicznym oraz wymogami stawianymi badanym danym wejściowym. W związku z tym, istnieje potrzeba klasyfikacji metod używanych w analizie sentymantu. Istnieje wiele sposobów na klasyfikację metod w analizie danych. W tej pracy jednak skupię się na klasyfikacji podejścia w analizie sentymantów opisanej przez Annę Baj-Rogowską<sup>29</sup>, która wyszczególniła 3 podejścia: metody słownikowe, metody uczenia maszynowego (podzielone na nadzorowane i nienadzorowane uczenie maszynowe) oraz podejście hybrydowe. Metody uczenia maszynowego są też nazywane metodami statystycznymi przez Krzysztofa Tomanka.<sup>30</sup>

Pierwszym omawianym rodzajem podejścia w analizie sentymantu według Anny Baj-Rogowskiej jest podejście hybrydowe. Jak nazwa wskazuje, to podejście polega na łączeniu elementów z metod słownikowych z metodami uczenia maszynowego, co pozwala na użycie szybkości metod słownikowych z precyzją metod uczenia maszynowego. Następnie przejdźmy do wyjaśnienia czym dokładnie są metody słownikowe oraz metody uczenia maszynowego.

---

<sup>27</sup> Cambria E., Schuller B., Xia Y., Havasi C., *New Avenues in Opinion Mining and Sentiment Analysis*, IEEE Computer Society, United States of America s. 16

<sup>28</sup> Liu B., *Sentiment Analysis and Opinion Mining...*, dz. cyt. s. 18

<sup>29</sup> Baj-Rogowska A., *Analiza sentymantu jako narzędzie monitorowania wyników finansowych przedsiębiorstwa, „Przedsiębiorczość i Zarządzanie”*, 2019, nr 20(12), s. 127

<sup>30</sup> Tomanek K., *Analiza sentymantu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych*, „*Przegląd Socjologii Jakościowej*”, 2014, nr 10(2), s. 121

### 2.3.1. Metody słownikowe

Metody słownikowe (leksykalne), jak sama nazwa tej metody oznajmia, polega na użyciu gotowych słowników, zawierających słowa oraz wyrażenia. Tym словам oraz wyrażeniom przypisano wartość sentymentu, która wskazuje na stopień emocjonalnego wydźwięku. Wartość sentymentu może być: pozytywna, neutralna lub negatywna<sup>31</sup>. Terminy z pozytywnym sentymentem zwiększą ocenę sentymentu, a z negatywnym zmniejszą ocenę sentymentu. Słowa o neutralnym sentymencie nie zmniejszą oceny sentymentu. Istnieją słowniki posiadające warunki logiczne oraz reguły leksykalne w celu określenia sentymentu słowa. Warunki logiczne można podzielić na warunki, według użytego w zdaniu operatora: jednoargumentowego (np. nie), wieloargumentowego (koniunkcja, alternatywa, alternatywa wykluczająca, implikacja, ekwiwalencja), kontekstowymi (np. blisko, nie blisko, przed, nie przed, po, nie po). Reguły leksykalne odnoszą się do obszarów: podobieństw i różnić znaczenia wyrazów (synonimy, homonimy, antonimy, hiperonimy, meronimy), zasobów słów danego języka, gramatyki języka (miejscie słów w zdaniu)<sup>32</sup>. Przykładem warunku logicznego kontekstu jest dodanie słowa „nie” przed słowem o wydźwięku pozytywnym. Daje to w wyniku, sentyment negatywny. Niektóre słowniki zostają rozszerzone o możliwość przypisania emocji podstawowych do słów, opisanych w pierwszym rozdziale. Omówimy teraz przykłady modeli podejścia słownikowego używanych w analizie sentymentu:

- VADER (Valence Aware Dictionary and sEntiment Reasoner) – jest narzędziem służącym do analizy sentymentu w mediach społecznościowych i w tekście nieformalnym. Sprawdza się dobrze w analizie sentymentu w krótkich wypowiedziach, jak komentarze i posty w mediach społecznościowych, ocenach produktów czy wypowiedziach posiadających slang, znaki emocji czy skróty myślowe. VADER przez analizę polarności (polarity) słów i przypisywanie im wartości sentymentu, która jest zsumowana dla wszystkich słów w całym tekście. VADER używa wyniku złożonego, który mieści się w zakresie: -1 do +1. Wynik złożony powyżej 0,05 oznacza sentyment pozytywny, wynik złożony poniżej -0,05 oznacza sentyment negatywny. Wynik złożony pomiędzy -0,05 a 0,05 oznacza sentyment neutralny<sup>33</sup>. VADER jest zawarty w bibliotece NLTK (Natural Language Toolkit).

---

<sup>31</sup> Analiza sentymentu, Encyklopedia Zarządzania, [https://mfles.pl/pl/index.php/Analiza\\_sentymentu](https://mfles.pl/pl/index.php/Analiza_sentymentu) [dostęp: 29.06.2025]

<sup>32</sup> Tomanek K., Analiza sentymentu – metoda analizy danych jakościowych. ..., dz. cyt., s. 121

<sup>33</sup> Analiza sentymentu przy użyciu VADER, <https://www.geeksforgeeks.org/python/python-sentiment-analysis-using-vader/> [dostęp: 29.06.2025]

- TextBlob – narzędzie do analizy sentymantu, pozwalające tak jak VADER analizować sentyment w tekście. Różni się od VADER tym, że analizuje tekst o dodatkowy wskaźnik, oprócz polarności, subiektywność (subjectivity). Wynik subiektywności mieści się pomiędzy 0,00, co oznacza pełną obiektywność, a 1,00, co oznacza pełną subiektywność. Subiektywność pozwala ocenić jak bardzo tekst jest zbudowany na emocjach autora tekstu<sup>34</sup>. Textblob jest zbudowany na podstawie bibliotek NLTK oraz Pattern.
- SentiWordNet – narzędzie do analizy sentymantu, które jest rozszerzeniem narzędzia WordNet. W tym narzędziu (tak jak w WordNet) rzeczowniki, czasowniki, przymiotniki i przysłówki są pogrupowane w zestawy synonimów kognitywnych zwanych synsetami. Różni się od WordNet tym, że każdy synset ma trzy wartości: pozytywną (PosScore), negatywną (NegScore) oraz obiektywną (ObjScore). Suma tych wartości dla synsetu wynosi 1. To narzędzie jest zbudowane do sprawdzania sentymantu dla pojedynczych słów,. Jeśli chcemy przeprowadzić analizę całego zdania trzeba wykonać tokenizację, lemityzację oraz dodać oznaczenia słów<sup>35</sup>.
- NRC Emotion Lexicon (EmoLex) – narzędzie służące do analizy sentymantu w wypowiedzi. Cechuje się tym, że oprócz analizy polarności, przypisuje do każdego słowa emocje według wcześniej omawianej klasyfikacji emocji Plutchika (radość, smutek, gniew, strach, zaskoczenie, przeczuwanie, zaufanie, wstręt)<sup>36</sup>. Słownik NRC zawiera 14,182 słów w języku angielskim.

### 2.3.2. Uczenie maszynowe

Metody uczenia maszynowego opierają się na metodach statystycznych. Można wyróżnić w nich dwa podejścia: uczenie nadzorowane (supervised learning) i nienadzorowane (unsupervised learning). Uczenie nadzorowane polega na wcześniejszym przygotowaniu danych, poprzez przypisanie oznaczeń wydźwięku emocjonalnego do każdej wypowiedzi przez adnotatora, który musi zbadać każdą wypowiedź, co jest czasochronnym zadaniem. Po przygotowaniu danych przez człowieka, algorytm uczy się rozpoznawać wzorce w podanym tekście. To podejście pozwala na osiągnięcie precyzji oceny sentymentów, kosztem dłuższej analizy danych na początku. W tym podejściu stosuje się takich technik jak drzewa decyzyjne, reguły decyzyjne oraz sieci neuronowe. Drugie podejście, uczenie nienadzorowane nie wymaga wcześniejszego oznaczenia danych. W tej metodzie dzieli się tekst na zdania, po czym dokonuje

---

<sup>34</sup> Textblob, <https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis> [dostęp:29.05.2025]

<sup>35</sup> Esuli A., Sebastiani F., *SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining*, Italy, 2006, s. 417

<sup>36</sup> Zad S, Jimenez J., Finlayson M., *Hell Hath No Fury? Correcting Bias in the NRC Emotion Lexicon*, Association for Computational Linguistics, 2021, s. 104

się klasyfikacji zdań za pomocą list słów kluczowych i obliczonych miar podobieństw. Podejście to wiąże się z mniejszą dokładnością niż uczenie nadzorowane, ale wymaga mniejszego zaangażowania ludzkiego w etapie początkowym. Technikami stosowanymi w tym podejściu są metody hierarchiczne oraz metody oparte na miarach podobieństwa (np. indeks Jaccarda):<sup>37</sup>

- Logistic Regression – jest to model regresji logistycznej metody statystycznej. stosowana do klasyfikacji binarnej, której celem jest przewidywanie prawdopodobieństwa wystąpienia zdarzenia lub wyniku. Ten algorytm jest często stosowany ze względu na swoją prostotę oraz łatwość interpretacji. Po wytrenowaniu modelu regresji logistycznej można go wykorzystać do prognozowania nowych danych poprzez obliczenie przewidywanych prawdopodobieństw dla każdej obserwacji i zastosowanie progu decyzyjnego (zwykle 0,5) w celu zaklasyfikowania obserwacji do odpowiednich klas.<sup>38</sup>
- Naive Bayes – technika klasyfikacyjna oparta na Twierdzeniu Bayesa. Jest uznawany jako najłatwiejszy i najsilniejszy algorytm uczenia maszynowego. Naive Bayes jest oparty na prawdopodobieństwu warunkowym i zakłada, że cechy używane do przewidywania klasy są od siebie niezależne<sup>39</sup>. Pozwala ocenić czy sentyment jest pozytywny, negatywny czy neutralny. W tym celu trzeba usunąć znaki specjalne i zmienić duże litery na małe, lematyzacje, tokenizacje i wektoryzacje i trenowanie modelu przed oceną dokładności modelu.<sup>40</sup>
- Random Forest – polega na tworzeniu wielu drzew decyzyjnych, z czego każde drzewo uczy się przy użyciu części danych wybranych losowo. W wyniku czego polarność jest zdecydowana na podstawie ilości drzew z największą ilością danej polarności, przykładowo jeśli większość drzew wskaże na polarność neutralną, to wynik polarności random forest będzie neutralny.<sup>41</sup>

---

<sup>37</sup> Tomanek K., *Analiza sentymentu – metoda analizy danych jakościowych. ...*, dz. cyt., s. 121

<sup>38</sup> Analiza sentymentu przy użyciu regresji logistycznej, <https://medium.com/@nirajan.acharya777/sentimental-analysis-using-linear-regression-86764bfde907> [dostęp 29.06.2025]

<sup>39</sup> Naive Bayes, <https://www.geeksforgeeks.org/machine-learning/naive-bayes-classifiers/> [dostęp: 29.06.2025]

<sup>40</sup> Analiza sentymentu przy użyciu Naive Bayes, <https://medium.com/@zubairashfaque/sentiment-analysis-with-naive-bayes-algorithm-a31021764fb4> [dostęp: 29.06.2025]

<sup>41</sup> Bahrawi B., *SENTIMENT ANALYSIS USING RANDOM FOREST ALGORITHM ONLINE SOCIAL MEDIA BASED*, “Journal of Information Technology and Its Utilization”, 2019, nr 2(2), s. 29-33

## ROZDZIAŁ 3

### 3.1. Opis badań i danych

Analiza został przeprowadzona na danych rzeczywistych z witryny Twitter, które zostały pobrane przy użyciu Twitter API. Dane zostały pobrane z platformy Kaggle.

Zbiór danych obejmuje 1,6 miliona postów z platformy Twitter. Ponieważ nie potrzebowałem tyle danych na wykonaną tutaj analizę, postanowiłem się ograniczyć do 2000 postów. Każdy z postów w tym zbiorze ma przypisane swój własny nr identyfikacyjny. Polaryzacja w kolumnie ‘target’ została wykonana przy użyciu distant supervision w uczeniu maszynowym, które łączy elementy uczenia maszynowego nadzorowanego z nienadzorowanym. Wszystkie posty zostały ściągnięte z tylko pięciu okresów:

- Piątek 05 czerwca 11:05:33 2009
- Poniedziałek 15 czerwca 12:53:14 2009
- Poniedziałek 15 czerwca 13:39:50 2009
- Piątek 22 maja 05:10:17 2009
- Piątek 29 maja 13:40:04 2009

Wszystkie przesłane posty mają treść od 1 znaku minimum do 140 znaków, co było limitem znaków w czasie kiedy te posty zostały pobrane. Poza tym każdy wiersz w zbiorze danych posiada te informacje:

**Tabela 1 Opis zmiennych w kolumnach**

Nazwa kolumny	Opis
<b>target</b>	polaryzacja tweeta (0 = negatywna, 2 = neutralna, 4 = pozytywna), uzyskana automatycznie przy użyciu distant supervision.
<b>ids</b>	Identyfikator tweeta (np. 2087).
<b>date</b>	data tweeta (Sat May 16 23:58:44 UTC 2009).
<b>flag</b>	zapytanie (lyx). Jeśli nie ma zapytania, ta wartość to NO_QUERY.
<b>user</b>	użytkownik, który tweetował (np. robotickilldozr).
<b>text</b>	tekst tweeta (np. Lyx is cool).

Źródło: Opracowanie własne.

Celem poniższych badań jest porównawcza analiza dwóch wybranych metod wykorzystywanych do analizy emocji i wydźwięku komunikatów w mediach społecznościowych. Porównanie opiera się na: klasyfikacji sentymantu zastosowanej w zbiorze danych Sentiment140, opartej na automatycznych oznaczeniach opartych na emotikonach (distant supervision) oraz alternatywnym podejściu słownikowym z wykorzystaniem NRC Emotion Lexicon, które przypisuje emocje i wydźwięk na podstawie obecności konkretnych słów. Dokładnym celem badania jest ocena spójności wyników uzyskanych obiema metodami i identyfikacja różnic w rozpoznawaniu emocji i wydźwięku.

Badanie to pozwala lepiej zrozumieć, w jakim stopniu różne metody analizy emocji mogą być wykorzystywane w praktyce, np. w badaniach opinii publicznej, analizie zachowań konsumenckich, marketingu czy komunikacji kryzysowej.

### **3.2. Badania**

Przed rozpoczęciem mojego badania ograniczyłem się do 2 tysięcy postów z 1,6 miliona. Jest to uwarunkowane tym, że sprzęt na którym została wykonana analiza emocji, nie byłby w stanie ocenić wszystkich emocji w tekście w dostatecznym czasie. W związku z tym że w zbiorze danych jest dokładnie tyle samo wierszy o polaryzacji pozytywnej co i negatywnej zdecydowałem się stworzyć dwa podzbiory, jeden dla pozytywnej polaryzacji, drugi dla negatywnej polaryzacji z 1000 postem. Następnie połączylem te dwa zbiory.

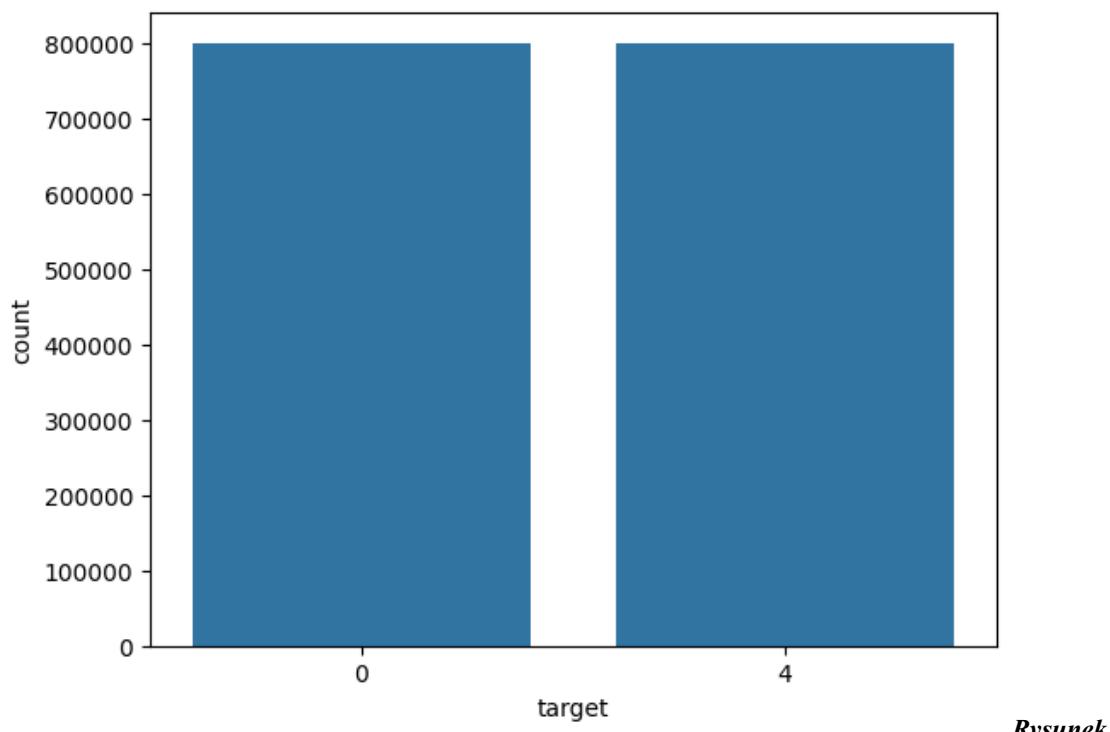
Mając już gotowy zbiór do pracy, rozpoczęłem czyszczenie danych. W pierwszym kroku wszystkie duże litery zostały zmienione na małe litery. Następnie usunąłem wszystkie znaki interpunkcyjne wraz z linkami internetowymi oraz znakiem specjalnym '@. Te zmiany zostały wykonane, ze względu na to że mogą one zniekształcić wynik analizy emocji.

Następnie usunięto słowa stopu i wykonano tokenizację. Słowa stopu zostały usunięte, ponieważ zakłócają one analizę emocji, tak jak znaki interpunkcyjne. Usunięcie ich także przyśpiesza analizę emocji. Wykonano tokenizację, ponieważ pozwala ona przygotować do lematyzacji i komputer lepiej opisuje pojedyncze tokeny, nie zbyt tekst.

Ostatecznie wykonano lematyzację. Lematyzacja ujednolica różne formy tego samego słowa, dzięki czemu możemy łatwiej liczyć i porównywać słowa. Lematyzacja zmniejsza też wymiar danych w analizie; wymiar danych będący liczbą cech lub zmiennych, na których operujemy. Opiera się ona na regułach gramatycznych i słownikowych. W wyniku lematyzacji mamy dane gotowe do analizy emocji i sentymentów.

### 3.2.1. Podział sentymentów

Chociaż w zmiennej ‘target’ jest zapisana informacja, o możliwej polaryzacji neutralnej treści w postach, to zdecydowałem się to sprawdzić dla pewności. Informacja o tym czy są posty o polaryzacji neutralnej, mogła wpływać negatywnie na początkowe analizy emocji oparte o słownik NRC Emolex, który nie wykrywa wbudowanej funkcji oznakowania sentymentu neutralnego. Jak się okazało, pomimo że opisie zbioru danych była informacja o polaryzacji neutralnej, po sprawdzeniu tego przy użyciu ‘sns.countplot’, nie znaleziono żadnych postów o polarności neutralnej.



Rysunek

#### 2 Podział sentymentów według uczenia maszynowego

Źródło: Opracowanie własne.

Jednak z ciekawości jak byłoby możliwe znalezienie tekstów o neutralnym sentymencie, dodałem 'funkcję 'label\_polarity'', która zwróci sentymenit neutralny, jeśli liczba sentymenitów pozytywnych i negatywnych w częściach tekstu w tekście pojedynczego posta jest równa:

**Tabela 2 Suma sentymenitów według NRC**

NRC sentymenit	Suma
Neutralnych	928
Pozytywnych	633
Negatywnych	439

Źródło: Opracowanie własne.

Jak widać w powyższej tabeli, dodana przeze mnie funkcja przypisała 928 z 2000 postów, sentymenit neutralny. Oczywiście sentymenit osiągnięty przez metodę słownikową NRC Emolex, nie jest tym samym co polarność, będący wynikiem klasyfikatora lub regresji. w uczeniu maszynowym osiągnięta poprzez uczenie się na podstawie cech tekstu. Sentyment w NRC jest oparty na przypisaniu emocji ze słownika, opisanego przez człowieka.

### **3.2.2. Frekwencja emocji i sentymenitów**

Frekwencja sentymenitów może być pokazana na jednej tabeli lub wykresie, ze względu na to jak słownik NRC jest zbudowany. Niektóre słów ma tylko etykiety emocji, niektóre mają tylko etykiety sentymenitów. Istnieją też słowa, które mają jednocześnie sentymenit i emocje. Frekwencja to proporcja przypisanych emocji i sentymenitów w całkowitej liczbie słów.

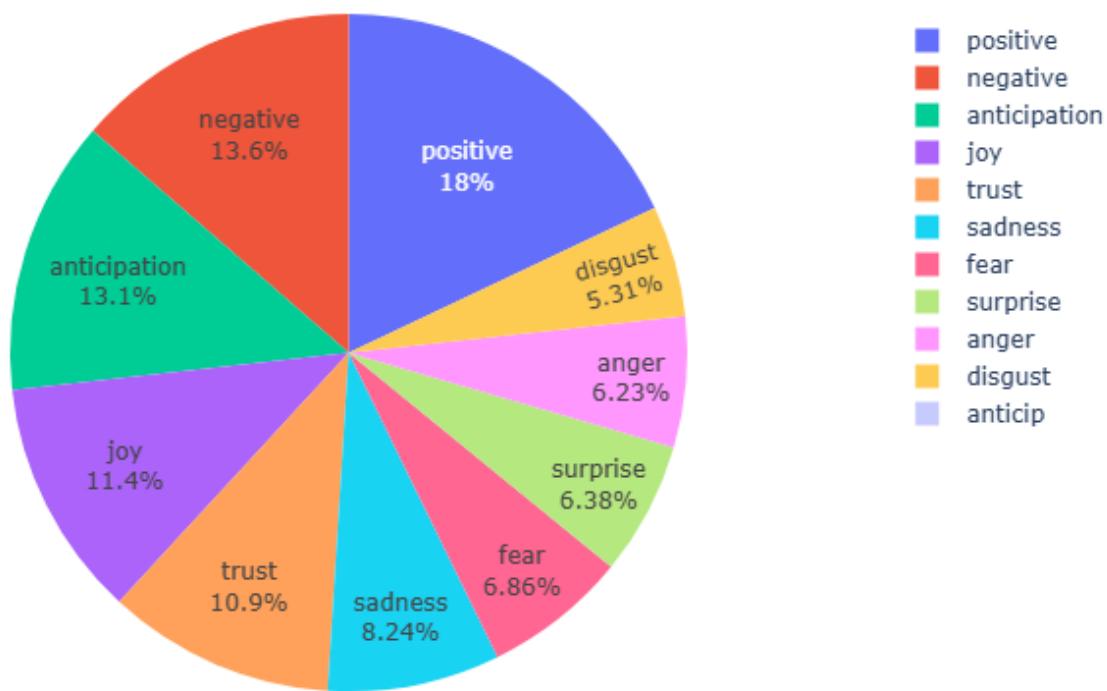
**Tabela 3 Frekwencja emocji i sentymenitów**

Emocje i sentymenity	Frekwencja
Pozytywne	0.179653 ~ 18%
Negatywne	0.136396 ~ 14%
Przeczuwanie	0.131182 ~ 13%
Radość	0.113851 ~ 11%
Zaufanie	0.108637 ~ 11%
Smutek	0.082429 ~ 8%
Strach	0.068621 ~ 7%
Zaskoczenie	0.063830 ~ 6%
Gniew	0.062280 ~ 6%
Wstręt	0.053121 ~ 5%

Źródło: Opracowanie własne.

Frekwencji wykazała, że sentyment pozytywny ma przewagę w z badanym tekście. Frekwencja sentymentu pozytywnego wynosiła 18%, w porównaniu do 14% sentymentu negatywnego. Wskazuje to na umiarkowaną obecność negatywnych tonów w analizowanych wypowiedziach. Interesujące jest wysokie występowanie przeczuwania, które stanowi 13%. Może się to wiązać z oczekiwaniem lub przewidywaniem, co świadczy o treściach skupionych na nadzieje lub obawach. Radość jest dość powszechna i stanowi około 11%, co zgadza się z wysokim poziomem emocji pozytywnych w porównaniu do negatywnych. Zaufanie również ma tą wartość, co może być związane z tonem wypowiedzi nacechowanym pewnością do innych. Emocje takie jak smutek (8%), strach (7%), zaskoczenie (7%), gniew (6%) i wstręt (5%) występują rzadziej, co może znaczyć, że mniej tekstów zawiera silne reakcje o takim charakterze.

### Frekwencja emocji dla tekstu

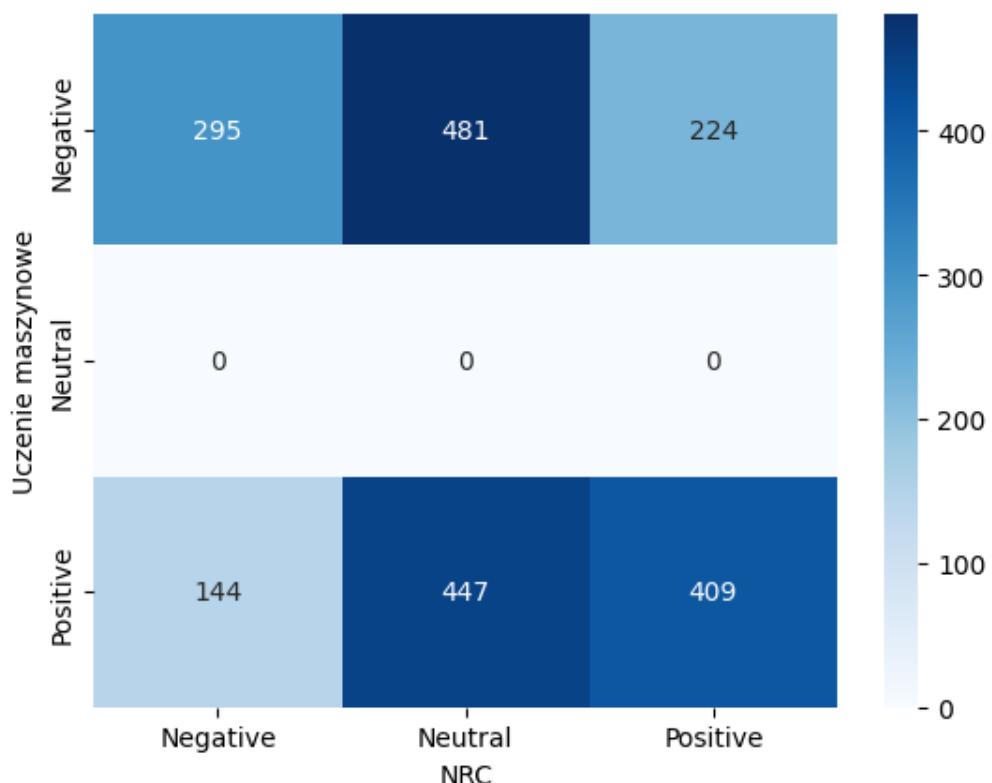


Rysunek 3 Frekwencja emocji dla tekstu

Źródło: Opracowanie własne.

### 3.2.3. Macierz sentymentów w NRC vs polarność uczeniu maszynowym

Jak było to wcześniej powiedziane, sentyment w NRC nie jest dokładnie tym co polarność w uczeniu maszynowym. Polarność to wynik klasyfikatora lub regresora, który na podstawie cech uczy się rozpoznawać sentyment, a sentyment jest oparty na przypisaniu słowom emocji na podstawie słownika i ich zliczeniu. Oboje jednak się odnoszą do emocjonalnej oceny tekstu, więc dobrze byłoby ocenić w jakim stopniu się nakładają:



Rysunek 4 Macierz sentymentów w NRC vs uczeniu maszynowym

Źródło: Opracowanie własne.

Analizując powyższy macierz, widoczne jest to że nie ma żadnej zgodność w sprawie polaryzacji neutralnej, której w badanym zbiorze nie było. Pomijając ten fragment, można zobaczyć wysoką zgodność w sprawie polarność negatywnej (481 z 500) i pozytywnej (447 z 500) dla uczenia maszynowego z stworzonym na potrzeby tej pracy sentymentem neutralnym. Również jest wysoka zgodność dla naturalnie będącego częścią NRC sentymentu

pozytywnego z polarnością pozytywną (409 z 500). Inne wartość nie przekraczają 300 punktów.

### 3.2.4. Rozkład emocji i sentymentów w zbiorze danych

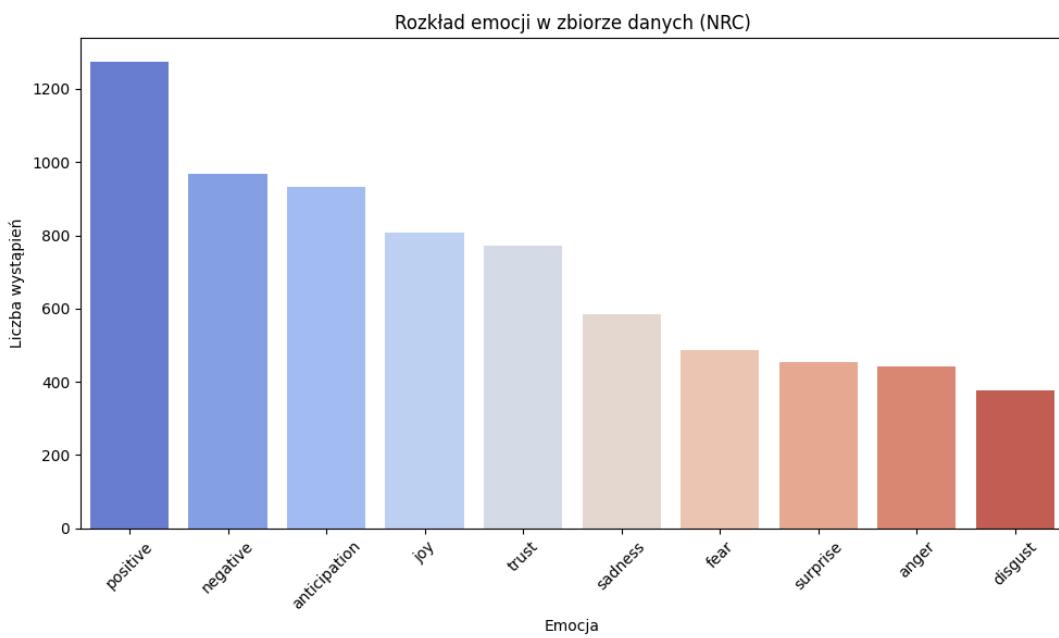
Poniżej przedstawiono dwa uzupełniające się wykresy dotyczące rozkładu emocji w zbiorze danych. Pierwszy z nich prezentuje ogólnączęstość występowania poszczególnych emocji, niezależnie od nadawcy, i pozwala określić, które emocje dominowały globalnie w treści tekstu. Drugi wykres skupia się na emocjach z perspektywy użytkowników – pokazuje, jaką emocję była dominująca w wypowiedziach każdego autora. Można więc powiedzieć, że pierwszy wykres odpowiada na pytanie: „Jak często występowała dana emocja w całym korpusie?”, a drugi zaś na pytanie: „Jaka emocja przeważała u poszczególnych użytkowników?”. Do lepszego zrozumienia pierwszego wykresu jest dołączona tabela:

Tabela 4 Rozkład emocji i sentymentów

Emocje i sentymenty	Częstość
Pozytywne	1275
Negatywne	968
Przeczuwanie	931
Radość	808
Zaufanie	771
Smutek	585
Strach	487
Zaskoczenie	453
Gniew	442
Wstręt	377

Źródło: Opracowanie własne.

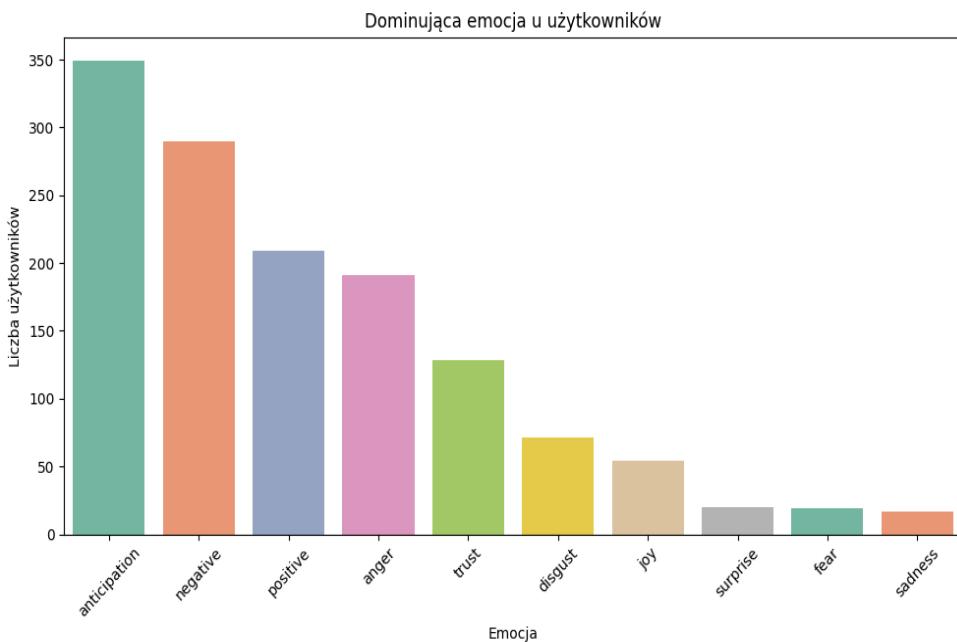
Pozytywne sentymenty występuje najczęściej w analizowanych, powtarzając się 1275, co oznacza, że ogólny ton wypowiedzi w tym zbiorze ma przewagę pozytywnego nastroju nad sentymentem negatywnym. Wysoka częstość radości (808) i zaufania (771) to potwierdza. Negatywne emocje są rozwożone równą pomiędzy sobą, nie występują różnice częstości pomiędzy nimi większe niż 250 użyć. Przeczuwanie (931) zajmuje trzecie miejsce pod względem częstości, co może wskazywać na duży ładunek emocjonalnego oczekiwania lub napięcia w wypowiedziach. Jest to emocja pośrednia, pomiędzy negatywnym, a pozytywnym tonem, w zależności od kontekstu.



**Rysunek 5 Rozkład emocji w zbiorze danych**

Źródło: Opracowanie własne.

Poniższy wykres przedstawia, które emocje były dominujące w wypowiedziach każdego autora:



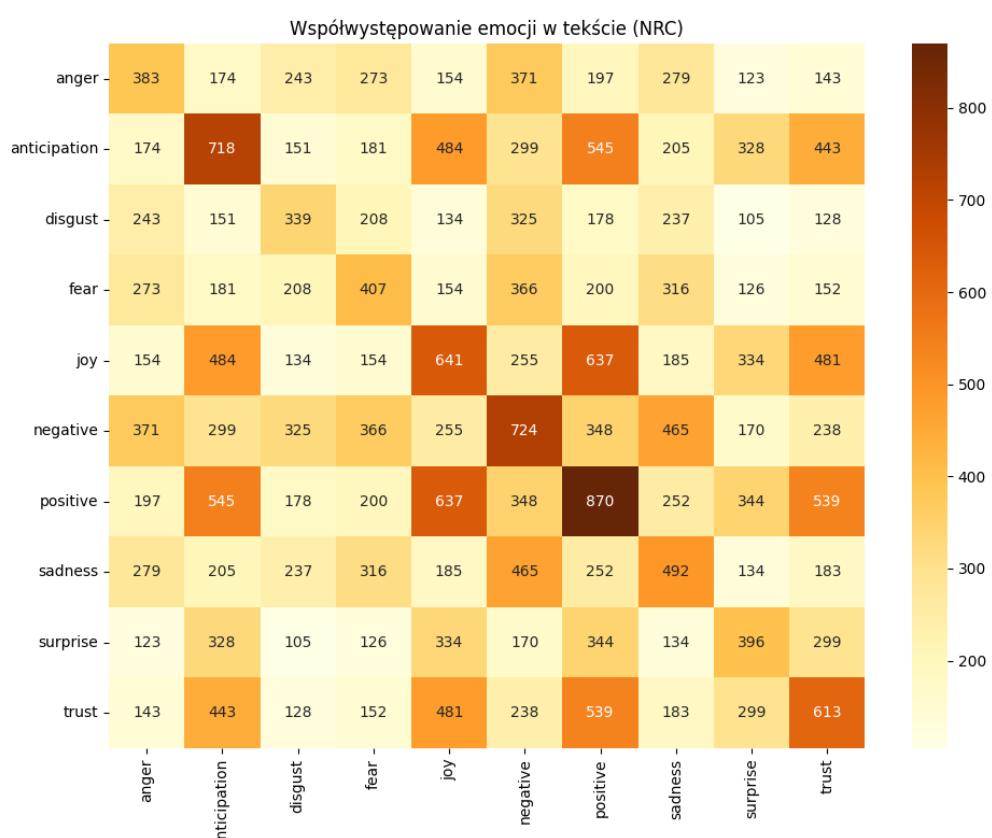
**Rysunek 6 Dominujące emocje u użytkowników**

Źródło: Opracowanie własne.

U poszczególnych autorów tekstu przeważała emocja przeczuwania, było tej emocji w inwidualnych autorów więcej niż sentymentów negatywnych i pozytywnych. Pozostałe emocje o tonie pozytywnym i negatywnym pojawiły się w mniejszych ilościach na przemian. Autorów tekstu, gdzie dominowały emocje zaskoczenia, strachu i smutku było mniej niż 50 na każdą z wspomnianych emocji.

### 3.2.5. Macierz współwystępowania

Poniżej znajduję się macierz współwystępowania emocji i sentymentów w tekście. Celem tej analizy jest uchwycenie zależności pomiędzy emocjami — czyli zbadanie, które z nich najczęściej pojawiają się razem w tej samej treści. Dzięki temu możliwe jest lepsze zrozumienie struktury emocjonalnej tekstów oraz identyfikacja najczęściej towarzyszących sobie stanów emocjonalnych.



Rysunek 7 Macierz współwystępowania w tekście (NRC)

Źródło: Opracowanie własne.

Najśilniejsze współwystępowanie istnieje pomiędzy sentymentem pozytywnym i radością (637), drugie pomiędzy najsilniejsze pozytywne współwystępowanie między sentymentem pozytywnym a zaufaniem (539). Emocje przeczuwania i sentymentu pozytywnego (545) także mają silne współwystępowanie.

Emocje jak przeczuwanie czy sentyment pozytywny mają wysokie wyniki współwystępowania z wieloma innymi emocjami, co mówi o ich roli łącznika w wypowiedziach. Macierz współwystępowania emocji wskazuje, że emocje w tekstu rzadko występują w izolacji. Często tworzą zestawy, które odpowiadają naturalnym emocjonalnym reakcjom człowieka – np. strach ze smutkiem, radość z zaufaniem, złość z odrazą.

## ZAKOŃCZENIE

Analiza emocji i sentymentów w badanym zbiorze danych pokazała kilka ciekawych rzeczy. Przede wszystkim – mimo że analizowany był tylko niewielki fragment (2000 postów z 1,6 miliona), udało się zauważać wyraźną przewagę emocji pozytywnych nad negatywnymi. Najczęściej występował sentyment pozytywny, a zaraz za nim takie emocje jak radość i zaufanie. Pokazuje to, że ogólny ton wypowiedzi w analizowanej próbce był raczej optymistyczny i przyjazny.

Zaskakującym wynikiem była wysoka obecność emocji określonej jako „przeczuwanie” – zarówno w całym zbiorze, jak i jako dominująca emocja u wielu użytkowników. Może to oznaczać, że wiele postów dotyczyło oczekiwania, przemyśleń o przyszłości albo wyrażania niepokoju i nadziei. Przeczuwanie wydaje się emocją „pośrednią” – może być zarówno pozytywne, jak i negatywne w zależności od kontekstu.

Jeśli chodzi o emocje negatywne, to były one bardziej rozproszone – nie dominowała żadna z nich wyraźnie. Gniew, strach, smutek czy wstręt pojawiały się w zbliżonych ilościach, co sugeruje większe zróżnicowanie nastrojów wśród wypowiedzi nacechowanych negatywnie.

Ciekawy obraz dała też macierz współwystępowania emocji, która pokazała, że emocje często pojawiają się razem – np. radość często łączyła się z pozytywnym sentymentem, a zaufanie z przeczuwaniem. Widać też powiązania między emocjami negatywnymi – np. smutek często występował ze strachem. Można powiedzieć, że emocje w tekście „chodzą parami” i rzadko występują całkowicie samodzielnie.

Porównanie wyników uzyskanych z użyciem słownika NRC i metod uczenia maszynowego pokazało, że w przypadku tekstów jednoznacznie pozytywnych lub negatywnych wyniki były bardzo zbliżone. W przypadku tekstów neutralnych ta zgodność już się nie pojawiła, co wynika z ograniczeń samego słownika NRC – nie ma on domyślnie kategorii neutralnej.

Podsumowując, analiza pokazała, że narzędzie NRC Emolex dobrze sprawdza się w ogólnym opisie emocjonalnego tonu wypowiedzi. Pozwala nie tylko policzyć, które emocje są najczęstsze, ale też sprawdzić, jak się ze sobą łączą i jak rozkładają się między użytkownikami. Dzięki temu można lepiej zrozumieć, co czuli autorzy tekstów i jak wyrażali swoje emocje.

## Bibliografia

### Pozycje zwarte

1. Doliński D., *Psychologia. Podręcznik akademicki*, t. 2, Gdańsk: Gdańskie Wydawnictwo Psychologiczne, 2004.
2. Kalat J. W., *Biologiczne podstawy psychologii*, Warszawa: Wydawnictwo Naukowe PWN, 2006.
3. Liu B., *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
4. Lula P., *Text mining jako narzędzie pozyskiwania informacji z dokumentów tekstowych*, Kraków: Statsoft, 2005.

### Artykuly naukowe

1. Hatfield G., “Did Descartes Have a Jamesian Theory of the Emotions?”, *Philosophical Psychology*, vol. 20, nr 4, 2007.
2. Krzemkowska-Saja J., “Namiętności duszy a współczesny spór o naturę emocji”, *Filosofia*, nr 17(2), 2012.
3. Judycki S., “Introspekcja jako problem filozoficzny”, *Roczniki Filozoficzne*, t. 50, z. 1, 2002.
4. Plutchik R., “Emotion: A psychoevolutionary synthesis”, *The American Journal of Psychology*, nr 93(4), 1980.
5. Plutchik R., “The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice”, *American Scientist*, nr 89(4), 2001.
6. Muchinsky P., “Emotions in the Workplace: The Neglect of Organizational Behavior”, *Journal of Organizational Behavior*, nr 21(7), 2000.
7. Lurie Y., “Humanizing Business through Emotions: On the Role of Emotions in Ethics”, *Journal of Business Ethics*, nr 49(1), 2004.
8. Solomon R., “The Moral Psychology of Business: Care and Compassion in the Corporation”, *Business Ethics Quarterly*, nr 8(3), 1998.

9. Holbrook M., Batra R., "Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising", *Journal of Consumer Research*, nr 14(3), 1987.
10. Hearst M., "Untangling Text Data Mining", *School of Information Management & Systems*, University of California, Berkeley, 1999.
11. Gładysz A., "Przegląd zastosowań analizy text miningowej", *Autobusy – Technika, Eksploatacja, Systemy Transportowe*, nr 12, 2016.
12. Tomanek K., "Metodyka dla analizy treści w projektach stosujących techniki text mining i rozwiązania CAQDAS piątej generacji", *Przegląd Socjologii Jakościowej*, nr 13(2), 2017.
13. Baj-Rogowska A., "Analiza sentymantu jako narzędzie monitorowania wyników finansowych przedsiębiorstwa", *Przedsiębiorczość i Zarządzanie*, nr 20(12), 2019.
14. Tomanek K., "Analiza sentymantu – metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych", *Przegląd Socjologii Jakościowej*, nr 10(2), 2014.
15. Cambria E., Schuller B., Xia Y., Havasi C., "New Avenues in Opinion Mining and Sentiment Analysis", *IEEE Computer Society*, USA.
16. Esuli A., Sebastiani F., "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining", Włochy, 2006.
17. Zad S., Jimenez J., Finlayson M., "Hell Hath No Fury? Correcting Bias in the NRC Emotion Lexicon", *Association for Computational Linguistics*, 2021.
18. Bahrawi B., "Sentiment Analysis Using Random Forest Algorithm Online Social Media Based", *Journal of Information Technology and Its Utilization*, nr 2(2), 2019.

## Strony internetowe

1. *Emocje*, Encyklopedia PWN, <https://encyklopedia.pwn.pl/szukaj/emocje> [dostęp: 05.02.2025].
2. *Analiza sentymantu*, Encyklopedia Zarządzania, [https://mfiles.pl/pl/index.php/Analiza\\_sentymentu](https://mfiles.pl/pl/index.php/Analiza_sentymentu) [dostęp: 29.06.2025].
3. *Analiza sentymantu przy użyciu VADER*, GeeksforGeeks, <https://www.geeksforgeeks.org/python/python-sentiment-analysis-using-vader/> [dostęp: 29.06.2025].

4. *TextBlob – Sentiment Analysis*,  
<https://textblob.readthedocs.io/en/dev/quickstart.html#sentiment-analysis> [dostęp: 29.05.2025].
5. *Sentiment Analysis Using Linear Regression*, Medium,  
<https://medium.com/@nirajan.acharya777/sentimental-analysis-using-linear-regression-86764bfde907> [dostęp: 29.06.2025].
6. *Naive Bayes*, GeeksforGeeks, <https://www.geeksforgeeks.org/machine-learning/naive-bayes-classifiers/> [dostęp: 29.06.2025].
7. *Sentiment Analysis with Naive Bayes Algorithm*, Medium,  
<https://medium.com/@zubairashfaque/sentiment-analysis-with-naive-bayes-algorithm-a31021764fb4> [dostęp: 29.06.2025].

## **SPIS TABEL**

**Tabela 5 Opis zmiennych w kolumnach**

**Tabela 6 Suma sentymentów według NRC**

**Tabela 7 Frekwencja emocji i sentymentów**

**Tabela 8 Rozkład emocji i sentymentów**

## **SPIS RYSUNKÓW**

Rysunek 1 Koło emocji Roberta Plutchika

Rysunek 4 Podział sentymentów według uczenia maszynowego

Rysunek 5 Frekwencja emocji dla tekstu

Rysunek 4 Macierz sentymentów w NRC vs uczeniu maszynowym

Rysunek 5 Rozkład emocji w zbiorze danych

Rysunek 6 Dominujące emocje u użytkowników

Rysunek 7 Macierz współwystępowania w tekście (NRC)

## **SPIS ZAŁĄCZNIKÓW**

Źródło danych: Sentiment140 dataset with 1.6 million tweets,  
<https://www.kaggle.com/datasets/kazanova/sentiment140> [dostęp: 29.06.2025].