**SPRAWOZDANIE**

Zajęcia: Analiza procesów uczenia

Prowadzący: prof. dr hab. inż. Vasyl Martsenyuk

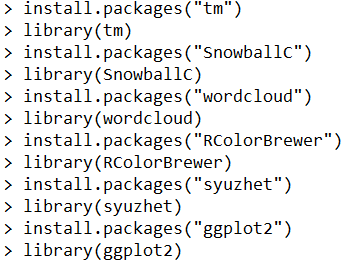
|  |  |
| --- | --- |
| Laboratorium Nr 7  Data 26.05.2023  Temat: Problemy NLP w uczeniu maszynowym  Wariant 4 | Maksymilian Grygiel  Informatyka  II stopień, stacjonarne,  Semestr I, gr.1a |

Link do repozytorium: <https://github.com/Maksiolo20/APU>

**Zadania:**

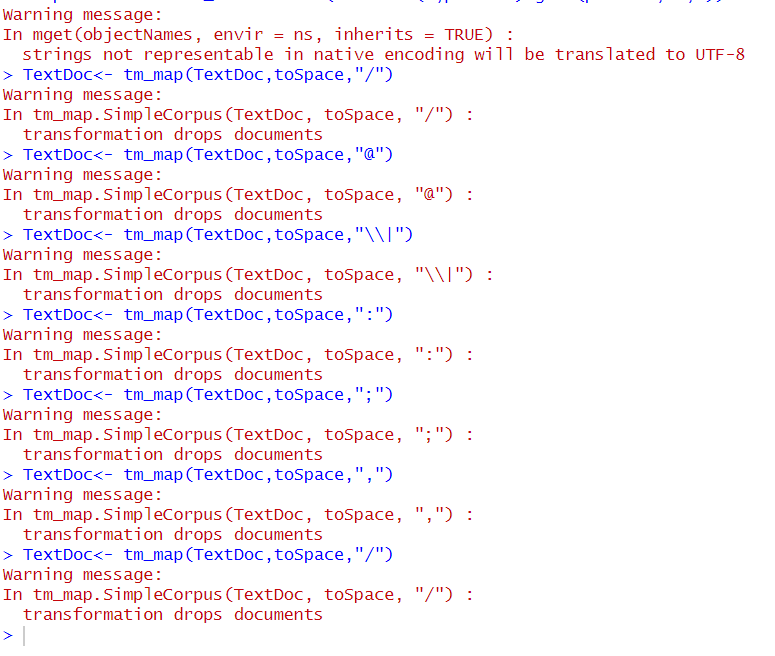
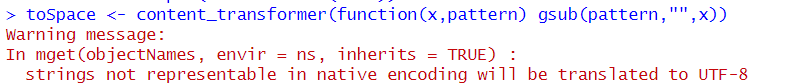
Zadanie dotyczy analizy tekstu, w tym listę częstotliwości słów, budowanie chmury słów, kojarzeń, sentiment analysis, emotion analysis, bigramów, grafów powiązań. Warianty zadania są określone tekstem w języku angielskim umieszczonym na portalu en.wikipedia.org (główna część artykułu **bez literatury**):  
https://en.wikipedia.org/wiki/History\_of\_poetry

**Wykonanie zadania:**

Instalacja pakietów:  


Odczytanie tekstu:  


Konwersja tekstu do obiektu:  


Czyszczenie tekstu ze zbędnych znaków:  
Uswanie specjalnych znaków:  


Usuwanie liczb:  
Inserting image...

Usuwanie znaków stop:  
Inserting image...

Usuwanie tajnych znaków:  
Inserting image...

Usuwanie znaków interpunkcyjnych:  
Inserting image...

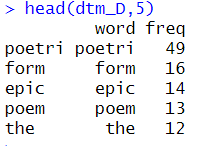
Usuwanie białych przestrzeniL  
Inserting image...

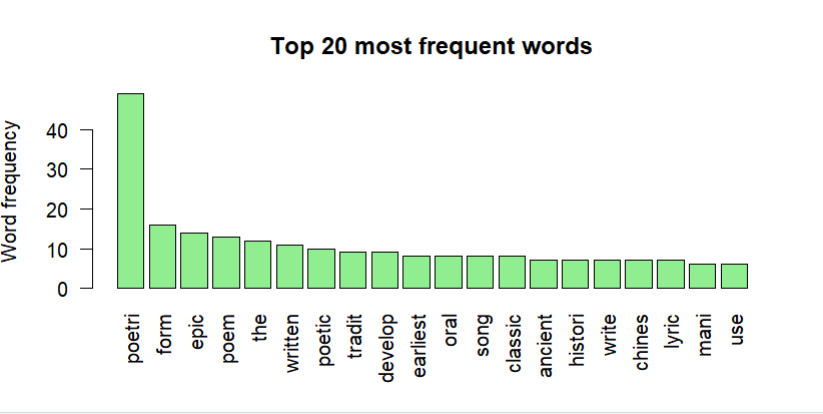
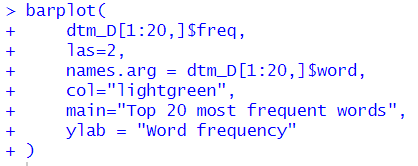
Zmiana do formy bazowej:  
Inserting image...

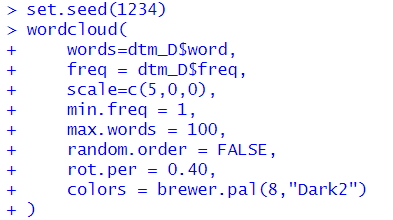
Budowanie macierzy tekstowej:  

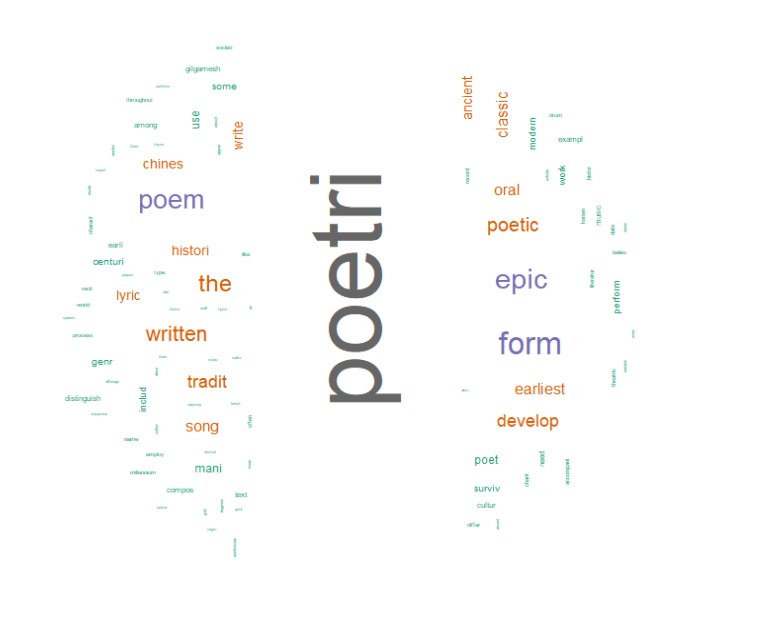

Sortowanie malejąco bazując na tym, jak częst słowo się pojawia:  

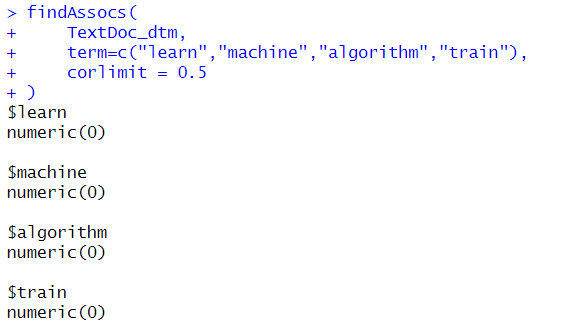

Pokaż 5 najczęstszych słów:

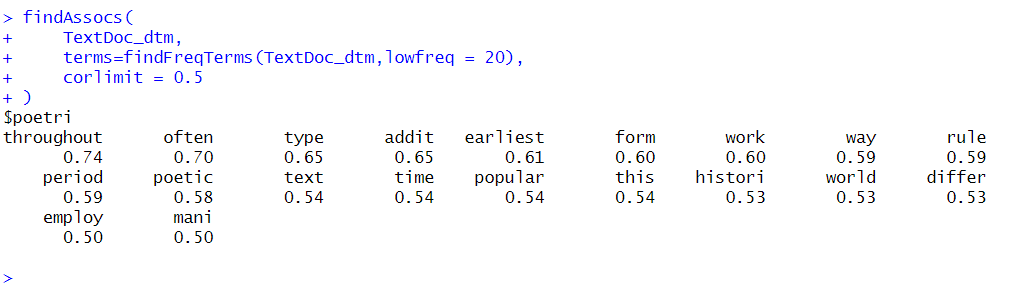


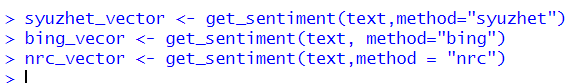
Najczęstsze słowa - wykres:  


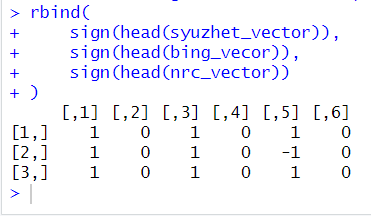
Wygenerowanie chmury słów:  


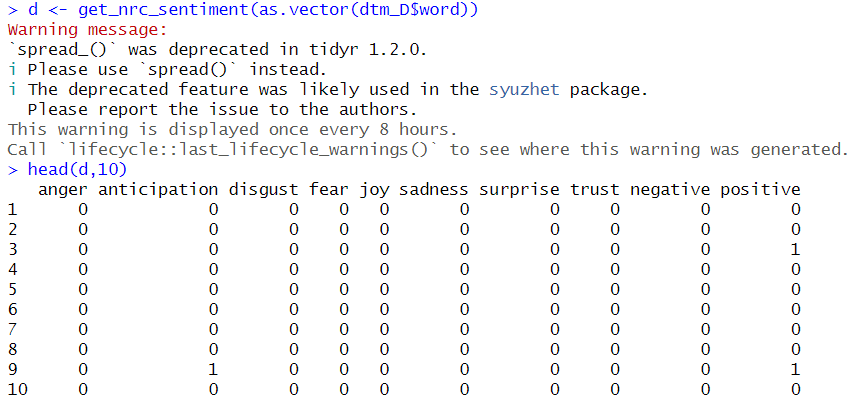


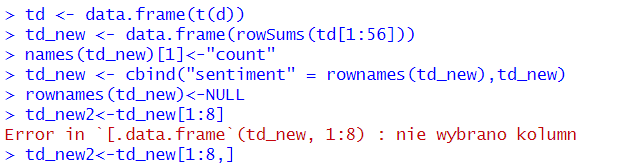
Kojarzenia słów:  


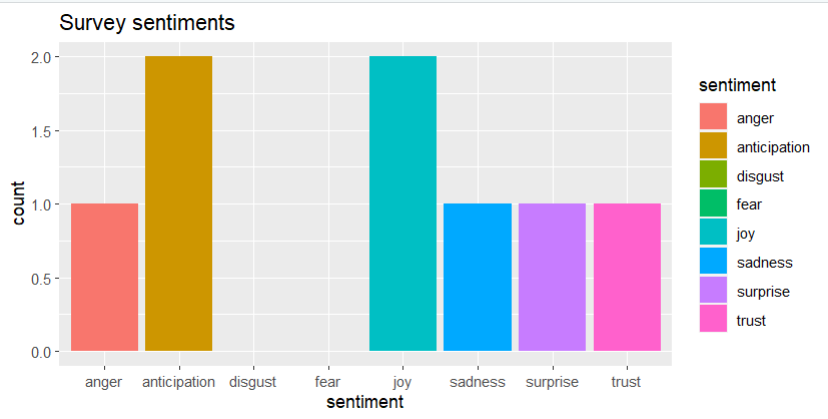
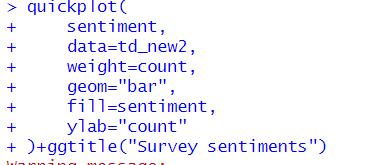
Znalezienie kojarzeń słów które pojawiają się co najmniej 20 razy:  


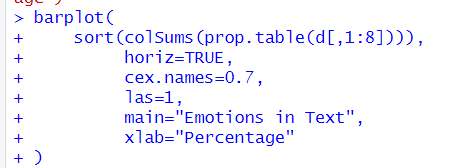
Analiza sentymentu:  


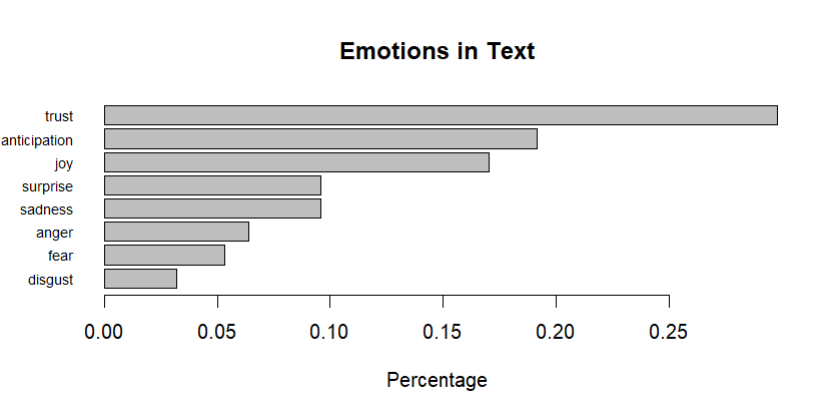
Porównanie analizy:  


Klasyfikacja emocji w tekście:  


Transpozycja, czysczenie wyników, sumowanie częstotliwości emocji w pierwszych 56 słowach:  


Drugi wykres - słowa przekazujące emocje:  


Trzeci wykres – procent emocji:  




**Wnioski:**

Sprawozdanie z R Studio na temat problemów związanych z przetwarzaniem języka naturalnego (NLP) w uczeniu maszynowym wykazało, że NLP jest obszarem o unikalnych wyzwaniach, takich jak zrozumienie i generowanie tekstu, analiza sentymentu oraz tłumaczenie maszynowe. Wnioskiem z tego sprawozdania jest, że skuteczne stosowanie uczenia maszynowego w obszarze NLP wymaga zastosowania specjalistycznych technik i narzędzi, takich jak tokenizacja, wektoryzacja tekstu czy modele językowe. Ponadto, wykorzystanie gotowych zbiorów danych oraz dostęp do mocnych obliczeniowo środowisk, takich jak R Studio, jest kluczowe dla skutecznego modelowania NLP.