**Dokumentacja Specyfikacji Wymagań**

### **1. Wprowadzenie**

Dokument przygotowany przez nas opisuje wymagania do projektu, który analizuje opinie klientów o perfumach marki Prada na stronie internetowej perfumerii Sephora. Projekt wykonałyśmy w języku R z wykorzystaniem podejścia Reproducible Research. Dane zebrałyśmy w pliku CSV, w którym są opinie użytkowników.

### **2. Cele systemu**

Zależało nam, aby stworzyć proste narzędzie do analizy danych tekstowych, dzięki któremu użytkownik ma możliwość:

* wczytania pliku z opiniami,
* przetworzenia i oczyszczenia tekstów,
* wykonania analizy słów i sentymentu,
* obejrzenia wyników w formie wykresów i chmury słów,
* wygenerowania raportu HTML z wynikami.

### **3. Wymagania funkcjonalne**

System powinien umożliwiać:

* wczytanie danych z pliku CSV,
* oczyszczenie tekstu z niepotrzebnych znaków,
* policzenie częstości występowania słów,
* analizę sentymentu (pozytywne/negatywne/neutralne),
* tworzenie wykresów słupkowych i chmury słów,
* zapis wyników do pliku HTML.

### **4. Wymagania niefunkcjonalne**

* projekt powinien być powtarzalny (ten sam wynik przy ponownym uruchomieniu),
* kod powinien być czytelny i podzielony na logiczne sekcje,
* projekt powinien działać na każdym komputerze z zainstalowanym R,
* analiza powinna być szybka przy pliku do kilku tysięcy opinii.

### **5. Interfejsy użytkownika i dane**

* użytkownik uruchamia analizę za pomocą pliku .Rmd,
* dane wejściowe to plik CSV zawierający kolumnę z opiniami tekstowymi,
* raport wynikowy jest generowany automatycznie w formacie HTML.

### **6. Słownictwo użyte w projekcie**

* opinia – tekstowy komentarz klienta lub recenzja produktu,
* chmura słów – graficzne przedstawienie, które słowa pojawiają się najczęściej,
* analiza sentymentu – sprawdzenie czy opinia jest pozytywna, negatywna czy neutralna,
* TF-IDF – miara określająca ważność słowa w tekście. Jest to sposób, aby zobaczyć, które słowa są naprawdę ważne, a nie tylko często powtarzalne
* Reproducible Research – pisanie kodu i raportu tak, żebt ktoś inny bez problemu mógł odtworzyć całą analizę.

**7. Przypadki użycia(Use Cases)**

UC1: Wykonanie analizy opinii  
Kto? użytkownik (np. student)  
Co robi? uruchamia plik .Rmd, który przetwarza dane i generuje raport  
Co trzeba mieć wcześniej? plik CSV z opiniami  
Efekt: użytkownik otrzymuje raport HTML z analizą

UC2: Przegląd wyników  
Kto? użytkownik  
Co robi? otwiera raport HTML, przegląda wykresy i wnioski  
Co trzeba wcześniej? analiza musi zostać zakończona  
Efekt: użytkownik widzi najczęściej występujące słowa oraz rozkład sentymentu

### **8. Scenariusze użytkownika (User Stories)**

* Jako student analizujący dane tekstowe chcę wczytać plik z opiniami o perfumach, żeby przeprowadzić analizę i przygotować raport na zaliczenie.
* Jako użytkownik narzędzia w R chcę automatycznie wygenerować wykresy i chmurę słów, żeby szybko zrozumieć, jakie cechy perfum są najczęściej komentowane przez klientów.
* Jako osoba tworząca raport Reproducible Research chcę mieć wszystkie kroki analizy zapisane w jednym pliku .Rmd, żeby wszystko było przejrzyste i można było łatwo odtworzyć i poprawić projekt w razie potrzeby.