

Specyfikacja Wymagań Systemowych

Projekt: Klasyfikacja - Recenzja E-sklepów z Ubraniami

Autor: Julia Godlewska

Data: 01.06.2025

Spis treści:

Specyfikacja Wymagań Systemowych	1
1. Wprowadzenie	1
2. Cele Systemu	1
3. Wymagania funkcjonalne	2
4. Wymagania нефunkcjonalne	2
5. Interfejsy użytkownika i wymagania dotyczące danych	2
6. Słownictwo dokumentacji	2
7. Przypadki użycia (use cases)	2
8. Scenariusze użytkownika (user stories)	3

1. Wprowadzenie

Niniejszy dokument stanowi Specyfikację Wymagań Systemowych (SRS) dla projektu "Klasyfikacja - Recenzja E-sklepów z Ubraniami". Dokument określa cele, wymagania funkcjonalne i нефunkcjonalne, interfejsy, słownictwo, przypadki użycia oraz scenariusze użytkownika dla systemu przeprowadzającego analizę eksploracyjną danych oraz trenującego model do klasyfikacji recenzji według kategorii binarnej.

2. Cele Systemu

Celem projektu jest stworzenie systemu umożliwiającego analizę recenzji produktów sklepów z ubraniami i predykcję ich negatywnego lub pozytywnego sentymentu. System powinien umożliwiać:

- Wprowadzanie danych tekstowych
- Przygotowanie danych do analizy
- Wykorzystanie modelu liniowego do predykcji kwalifikacji binarnej
- Wizualizację wyników za pomocą chmur słów i wykresów

3. Wymagania funkcjonalne

- Możliwość wprowadzania danych tekstowych
- Oczyszczanie i przetwarzanie danych
- Umożliwić analizę danych poprzez chmury słów
- Wprowadzić model predykcji na bazie losowego podziału danych
- Wprowadzić model predykcji na bazie stratyfikowanego podziału danych
- Umożliwić ocenę skuteczności modeli
- Wizualizacja wyników modeli w postaci wykresów

4. Wymagania niefunkcjonalne

- Szybkość analizy
- Czytelność wyników
- Czytelność kodu
- Łatwość modyfikacji systemu

5. Interfejsy użytkownika i wymagania dotyczące danych

- System uruchamiany w środowisku obsługującym kompilację kodu R
- Interfejs użytkownika ogranicza się do wprowadzania danych i dostosowania parametrów
- Dane wejściowe w postaci pliku csv zawierającego dane tekstowe

6. Słownictwo dokumentacji

- Chmura słów - graficzna reprezentacja częstości występowania słów w zbiorze tekstów
- Podział losowy - metoda podziału danych poprzez przypadkowe przydzielenie próbek
- Podział stratyfikowany - metoda podziału danych, która uwzględnia rozkład kategorii, aby zachować podobne proporcje w każdej z podgrup

7. Przypadki użycia (use cases)

1. **Analiza i wizualizacja danych:** System czyści tekst, generuje chmurę słów i listy częstości słów.
2. **Trenowanie modelu (podział losowy):** Dane dzielone są losowo (80/20), model jest trenowany i oceniany za pomocą macierzy pomyłek.
3. **Porównanie skuteczności modeli (losowy vs. stratyfikowany podział):** System testuje oba podejścia i przedstawia metryki porównawcze.

8. Scenariusze użytkownika (user stories)

1. **Jako właściciel sklepu**, chcę wygenerować chmurę słów i listy częstości występowania słów, aby zrozumieć główne tematy w pozytywnych i negatywnych recenzjach.
2. **Jako analityk danych**, chcę wytrenować model na losowo podzielonych danych (80% treningowe, 20% testowe), aby sprawdzić jego skuteczność.
3. **Jako kierownik projektu**, chcę porównać wyniki modeli przy losowym i stratyfikowanym podziale danych, aby wybrać bardziej wiarygodną metodę walidacji.