

Inżynieria Uczenia Maszynowego

Projekt: Etap 1

Dominika Boguszewska
Jędrzej Osiecki

Semestr 25L

1 Temat projektu

“Niektórzy goście oceniają zbyt surowo – czy będziemy w stanie ustalić którzy?”

2 Definicja problemu biznesowego

Celem projektu jest identyfikacja gości, którzy wystawiają zbyt surowe opinie. Owe opinie mogą najlepiej nie odzwierciedlać faktycznej jakości zakwaterowania, w wyniku czego mogą one zaniżać średnią ocenę lokalu. Identyfikacja ich może wspomóc zarządzanie relacjami z gośćmi i reputacją gospodarzy.

2.1 Przykładowe zastosowania modelu

- Automatyczne oznaczanie surowych opinii do przeglądu przez pracownika portalu Nocarz.
- Umożliwienie szybkiej reakcji gospodarzy na niesprawiedliwą krytykę.
- Usprawnienie procesu analizy opinii w różnych językach.

3 Zadanie modelowania

Klasyfikacja binarna – przewidywanie, czy dana opinia jest surowa czy nie.

3.1 Dane wejściowe

- Treść opinii (kolumna *comments*) – wielojęzyczna, wymaga przetwarzania języka naturalnego.
- Ocena numeryczna (kolumna *numerical_review*) – skala 0–5.
- Szczegółowe dane o lokalach - interesują nas głównie informacje o recenzjach, czyli kolumny *review_scores_rating* i *number_of_reviews*

3.2 Założenia

- Opinie surowe to silnie negatywnie nacechowane emocjonalnie oceny, które odbiegają od średniej ocen lokalu.
- Model powinien radzić sobie z opiniami w różnych językach.

4 Proponowane modele

4.1 Model bazowy

Modelem bazowym jest naiwny klasyfikator Bayesowski operujący na reprezentacji tekstu w postaci wag TF-IDF. Głównym założeniem tego modelu jest niezależność warunkowa cech, co upraszcza obliczenia kosztem precyzji przy bardziej złożonych związkach językowych.

Recenzje użytkowników, zapisane w różnych językach, zostaną przetłumaczone automatycznie na język angielski przy użyciu ogólnodostępnych usług tłumaczenia maszynowego (np. Google Translate).

Model ten zostanie wykorzystany jako punkt odniesienia dla późniejszej oceny skuteczności modelu docelowego. Pomimo swojej prostoty, Naive Bayes często osiąga zaskakująco dobre wyniki w zadaniach klasyfikacji tekstu, szczególnie przy małych zbiorach danych.

4.2 Model docelowy

Model docelowy to liniowy klasyfikator SVM (Support Vector Classifier). Kluczową zaletą tego modelu jest jego zdolność do znajdowania optymalnej hiperpłaszczyzny decyzyjnej w przestrzeni cech, co przekłada się na lepsze rozdzielenie klas niż w przypadku modeli probabilistycznych.

Model SVC wybrano jako model docelowy ze względu na jego zdolność do przetwarzania dużych zbiorów danych o wysokiej liczbie cech (typowe w klasyfikacji tekstu) oraz możliwość interpretacji wpływu poszczególnych cech na decyzję modelu poprzez analizę wag przypisanych słowom.

Przewiduje się, że model SVC osiągnie lepsze wyniki niż model bazowy dzięki bardziej elastycznej funkcji decyzyjnej i lepszemu dopasowaniu do danych.

Jako dane wejściowe zostaną wykorzystane recenzje użytkowników przetłumaczone na język angielski oraz numeryczna ocena noclegu.

5 Kryteria sukcesu

5.1 Analityczne kryterium sukcesu

Analityczne kryterium sukcesu opiera się na precyzyjnej ocenie skuteczności modelu klasyfikującego recenzje jako surowe. Sukces zostanie osiągnięty, jeżeli:

- model osiągnie wartość F1-score ≥ 0.70 dla klasy surowa,
- model osiągnie wartość precision ≥ 0.75 dla klasy surowa, co ograniczy liczbę fałszywych alarmów i zapewni, że większość recenzji oznaczonych przez model rzeczywiście zasługuje na uwagę moderatora.

Dodatkowo, model zostanie zweryfikowany poprzez analizę:

- fałszywych pozytywów – recenzji zaklasyfikowanych jako surowe, które według człowieka nie są problematyczne,
- fałszywych negatywów – recenzji niezaklasyfikowanych jako surowe, które w ocenie człowieka lub gospodarza miały jednak nadmiernie krytyczny ton.

Dzięki tym analizom możliwe będzie dostrojenie modelu w kolejnych iteracjach projektu oraz ocena jego praktycznej przydatności w kontekście rzeczywistego odbioru treści przez użytkowników.

5.2 Biznesowe kryterium sukcesu

Z punktu widzenia wartości biznesowej, system klasyfikujący recenzje powinien wspierać gospodarzy oraz moderatorów platformy w szybszym identyfikowaniu recenzji wymagających interwencji. Kryterium sukcesu obejmuje:

- $\geq 80\%$ recenzji zaklasyfikowanych jako „surowe” przez model powinno być uznanych również przez moderatorów lub gospodarzy jako faktycznie krytyczne lub potencjalnie szkodliwe,
- poprawa średniego czasu reakcji moderatorów na krytyczne opinie, dzięki wcześniejszemu sygnalizowaniu takich treści przez system,
- ograniczenie liczby przypadków, w których opinie o nadmiernie negatywnym wydźwięku pozostają niezauważone i bez odpowiedzi.

Dzięki automatyzacji detekcji surowych recenzji możliwe będzie bardziej proaktywne i spersonalizowane zarządzanie relacją z klientem, co w dłuższej perspektywie może wpłynąć pozytywnie na ocenę usług oraz wskaźnik ponownych rezerwacji.

6 Analiza danych

Zebrane dane są wystarczające do wstępnego trenowania modelu klasyfikującego recenzje jako „surowe” lub „niesurowe”, ale wymagają kilku przekształceń:

- recenzje należy przetłumaczyć na język angielski,
- należy zbalansować dane lub zastosować odpowiednie metody w przypadku klas nie zrównoważonych,
- warto wykorzystać długość recenzji oraz występowanie słów nacechowanych negatywnie jako cechy pomocnicze.

Dane sugerują również, że goście rzadko wystawiają bardzo niskie oceny, więc rzeczywiste „surowe” recenzje będą stanowiły niewielką część zbioru – co jest kluczowe dla konstrukcji i ewaluacji modelu.