Package 'shoppingwords'

July 23, 2025

Title Text Processing Tools for Turkish E-Commerce Data

Description Provides several datasets useful for processing and analysis of text in Turkish from an online shopping platform.

Version 0.1.0

Maintainer Betul Kan-Kilinc <bkan@eskisehir.edu.tr>

Imports stringi, stopwords, stringdist, tibble

Depends R (>= 4.0.0)

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

LazyData true

Suggests knitr, rmarkdown, dplyr, ggplot2

VignetteBuilder knitr

LazyDataCompression xz

URL https://bkanx.github.io/shoppingwords/

NeedsCompilation no

Author Betul Kan-Kilinc [aut, cre] (ORCID:
<https://orcid.org/0000-0002-3746-2327>),
Mine Çetinkaya-Rundel [ctb] (ORCID:
<https://orcid.org/0000-0001-6452-2420>),
Colin Rundel [ctb] (ORCID: <https://orcid.org/0000-0002-6058-8251>)

Repository CRAN

Date/Publication 2025-07-23 19:20:02 UTC

Contents

---

match_stopwords	*Remove Stopwords from User Reviews*

Description

This function processes a dataframe containing user reviews and removes predefined stopwords. It first searches the package's internal stopwords dataset (`stopwords_tr`), and if no match is found, it falls back to the broader `stopwords_iso` list.

Usage

```
match_stopwords(df)
```

Arguments

df   Dataframe containing user reviews, with required columns `comment` (text) and `rating` (numerical score).

Details

The function converts text to a standardized format by removing accents and special characters, transforming it into basic Latin characters, and making all letters lowercase. It then tokenizes the text, filters out stopwords, and returns the cleaned version.

Value

A modified dataframe with an additional `cleaned_text` column containing stopword-free text.

Examples

```
reviews_sample <- tibble::tibble(
  comment = c("Bu ürün xs ancak fiyatı yüksek gibi",
              "Fiyat çok pahalı ama kaliteli iyi"),
  rating = c(4.5, 3.0)
)
match_stopwords(reviews_sample)
```

---

phrases   *A dataset of phrases*

Description

Contains common negative-emotion phrases extracted from user reviews.

Usage

```
phrases
```

Format

A tbl_df with with 205 rows and 1 variable:

word ngrams.

Examples

```
phrases
```

---

reviews   *A dataset of reviews*

Description

User reviews collected from an e-commerce site.

Usage

```
reviews
```

Format

A tbl_df with with 260,308 rows and 3 variables:

rating Rating score, out of 5.

comment Comment text, in Turkish.

id Rating ID.

Examples

```
reviews
```

---

reviews_test *A test dataset*

Description

A test sample data used for testing analysis functions. It differs from reviews data. The `text` column in this data frame is similar to the `comment` column in the `reviews` data frame. Note that this data frame contains 170 `texts` that are in common, verbatim, with comments in the `reviews` dataset. This is because some users made the same comments. The `id` column shows that these are not the same observations, just similarly worded comments from different reviews.

Usage

```
reviews_test
```

Format

A tbl_df with with 1,481 rows and 4 variables:

rating Rating score, out of 5.

text Comment text, in Turkish.

emotion n for negative, p for positive.

id Rating ID.

Examples

```
reviews_test
```

---

stopwords_tr *A dataset of Turkish stopwords*

Description

A dataset of stopwords used in Turkish text analysis.

Usage

```
stopwords_tr
```

Format

A tbl_df with with 92 rows and 1 variable:

word Stopword, in Turkish.

Examples

```
stopwords_tr
```

Index