**Basic Library Text Preprocessing**

Dasar Text-Preprocessing dengan Python#

Case Folding: Lowercase

Case Folding: Removing Number

Case Folding: Removing Punctuation

Case Folding: Removing Whitespace

Separating Sentences with Split () Method

Tokenizing: Word Tokenizing Using NLTK Module

Tokenizing with Case Folding

Frequency Distribution

Frequency Distribution Visualization with Matplotlib

Tokenizing: Sentences Tokenizing Using NLTK Module

Filtering using NLTK

Filtering using Sastrawi: Stopword List

Filtering using Sastrawi

Add Custom Stopword

Stemming: Porter Stemming Algorithm using NLTK

Stemming Bahasa Indonesia using Sastrawi

**Text processing**

Cleaning Words

import library

remove tab new line and back slice

remove non ASCII emoticon chinese word

remove mention link hashtag

remove RT

lower

remove incomplete url

remove number

remove whitespace leading and trailing

remove multiple whitespace into single whitespace

remove single char

remove excessive dot

remove punctuation

remove excessive whitespace

**Preprocessing Eksprasis**

Menggunakan pembersihan kata dengan modul `Eksprasis` : https://github.com/cbaziotis/ekphrasis

**Stemming menggunakan Sastrawi**

Load dataset

Load sastrawi sebagai stemmer

Hapus stopwords

**Preprocessing nltk**

load library

load dataset

Preprocessing dataset

lakukan pembersihan dengan memanggil fungsi yang didefinisikan diatas

hapus data kosong

normalisasi kata slang menjadi kata baku

Simpan

**Preprocessing Menggunakan Spacy**

Perbedaan NLTK dan Spacy

Memakai NLTK

Memakai spacy

Load library

Load dataset

Preprocessing dataset

Lakukan pembersihan dengan memanggil fungsi yang didefinisikan diatas

Hapus data kosong

Normalisasi kata slang menjadi kata baku

Simpan