

# HASIL DATA MINING

## Terkait Penelitian

Ikang Fadhli

Nutrifood Indonesia

13 December 2021

## Section 1

# PENDAHULUAN

# Latar Belakang

Setelah *pilot project* dan diskusi yang lalu, berikutnya akan dicoba melakukan *data mining* kembali dengan menggunakan *keywords* yang berbeda dan lebih spesifik.

Pada kesempatan ini, saya akan kembali mencari **berbagai penelitian yang telah dilakukan di dalam negeri** terkait dengan beberapa *keywords* yang telah didefinisikan.

# Tujuan

Kali ini ini, saya mencoba untuk mencari berbagai penelitian terkait *keywords* berikut:

- *Indigenous Food* (termasuk padanan dalam bahasa Indonesianya: pangan lokal),
- *Functional Food*,
- *Fermented Food* (termasuk padanan dalam bahasa Indonesianya: makanan fermentasi),
- *Ethnic Food* (termasuk padanan dalam bahasa Indonesianya: makanan etnik),
- *Traditional Food* (termasuk padanan dalam bahasa Indonesianya: makanan tradisional),
- Makanan,
- Minuman

di situs [www.neliti.com](http://www.neliti.com) sebagai uji coba untuk melakukan analisa teks yang didapatkan. Dari hasil temuan yang ada, kita akan coba kembangkan *keywords* apa lagi yang mungkin akan muncul. Pada kesempatan mendatang, akan dilakukan *data mining* kembali untuk berbagai situs seperti *repository* perpustakaan berbagai universitas untuk mendapatkan gambaran penelitian yang telah dilakukan di universitas-universitas tersebut.

## Metode (*Data Mining / Web Scraping*)

Pengambilan data akan menggunakan algoritma *web scraping* dengan bahasa pemrograman **R** menggunakan *virtual machine* milik *Google Cloud*.

Situs yang dijadikan rujukan data adalah [www.neliti.com](http://www.neliti.com). Data yang akan diambil antara lain: judul penelitian dan *author* (termasuk *link* rujukan).

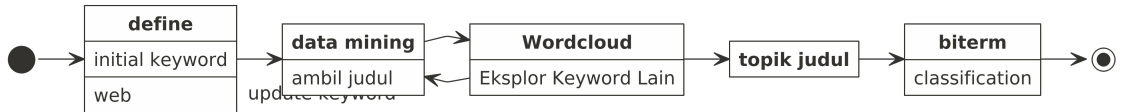
**Catatan penting:** hasil pencarian yang didapatkan murni berdasarkan output yang didapatkan dari situs *neliti*. Tidak ada jaminan bahwa semua penelitian tersebut selalu berkaitan penuh secara konten dengan *keywords* yang digunakan.

# Metode (*Text Analysis*)

Selanjutnya akan dilakukan beberapa *text analysis* seperti:

- ① *Word cloud*: untuk menemukan *keywords* lain yang mungkin berkaitan dengan *keywords* utama.
- ② *Biterm Topic Modelling*: untuk menentukan dan mengelompokkan judul artikel, penelitian, atau berita ke dalam topik-topik tertentu.

# Metode (lanjutan)



## Section 2

# HASIL DATA MINING

---

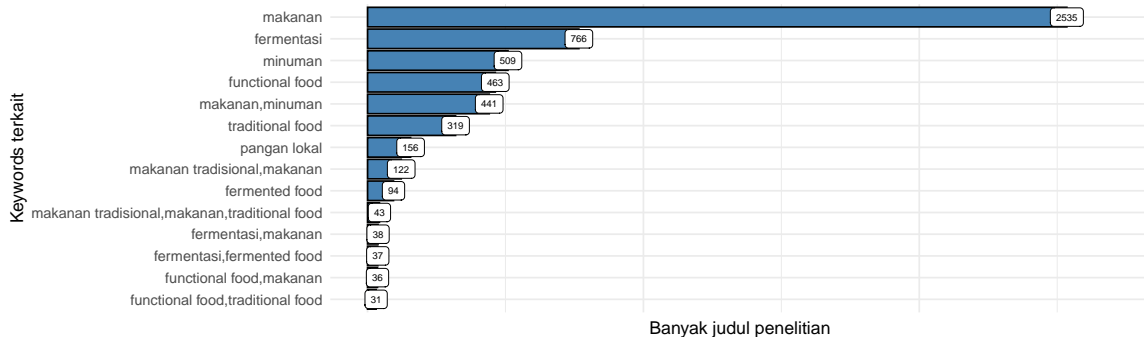


# Hasil Data Mining

Pada situs [www.neliti.com](http://www.neliti.com), didapatkan ada 12.841 buah *unique* penelitian hasil pencarian *keywords*. Tentunya bisa jadi satu judul penelitian keluar dari hasil pencarian lebih dari satu *keywords*. Berikut adalah grafik dari 14 *keywords* (dan kombinasi *keywords*) teratas berdasarkan banyaknya penelitian:

Berapa banyak penelitian yang didapatkan dari keywords ... ?

Hasil Data Mining Situs [www.neliti.com](http://www.neliti.com)



## Section 3

# ***TEXT ANALYSIS: Keywords Lain***

## Mencari *Keywords* Lain

Untuk mencari *keywords* lainnya, saya akan kumpulkan semua judul penelitian hasil pencarian lalu akan dihitung kata apa saja yang paling sering muncul.

Perlu diperhatikan bahwa kata sambung, kata depan, dan *stopwords* akan dihapus dari analisa ini.



# Mencari *Keywords* Lain

**lala**