



Data Article

Structured over-the-counter medicine dataset to improve the safety of pharmaceuticals



Danny Sebastian^{a,*}, Restyandito^a,
Justinus Putranto Agung Nugroho^b

^a *Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia*

^b *Fakultas Kedokteran, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia*

ARTICLE INFO

Article history:

Received 16 July 2024

Revised 19 July 2024

Accepted 29 July 2024

Available online 8 August 2024

Dataset link: [Pharma-safe Index Dataset](#)
([Original data](#))

Keywords:

Over-the-counter medicine dataset

Self-medication

Indonesian medicine dataset

Decision support system

Expert system

ABSTRACT

The Pharma-safe Index Dataset comprises comprehensive information on several aspects of pharmaceuticals, including general medication data, usage guidelines, dosage, adverse effects, pricing, drug interactions, duration of use, composition, and contraindications. The dataset is provided in both CSV and JSON file formats, and it is available in both English and Bahasa Indonesia. By conducting interviews, distributing flyers, and using pharmacy books, the dataset was collected from over-the-counter medications that were sold at three pharmacies located in Yogyakarta, Indonesia. A medical professional performed cleansing, standardization, and validation on it before it was exported to JSON and CSV formats. The data collected on drug efficacy, safety, and patient outcomes in Indonesia can be utilized by researchers in order to uncover trends and developing patterns of prescription drug resistance. It is possible that this will direct future research, lead to improvements in drug formulations, treatment strategies, and public health policies, and expand our understanding of how drugs work and how they affect patient health.

© 2024 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC license
(<http://creativecommons.org/licenses/by-nc/4.0/>)

* Corresponding author.

E-mail address: danny.sebastian@staff.ukdw.ac.id (D. Sebastian).

Specifications Table

Subject	Health Informatics, Data Mining and Statistical Analysis
Specific subject area	Public health; Pharmacology; Self-medicated; Decision Support System; Expert System
Type of data	Table
Data collection	Raw
Data source location	The dataset was collected from a list of over-the-counter drugs sold in three pharmacies located in the Yogyakarta Special Region, Indonesia. To complete the data, data collection was carried out through interviews with pharmacists and doctors, physical flyers from medicine boxes, and pharmacy books. Next, the data goes through a cleaning process and a structure standardization process using Microsoft Excel. The dataset results are validated by a doctor to guarantee the correctness of the data. Next the author and doctor translate the dataset from Indonesian to English. The final process of forming the dataset is to export it to JSON and CSV format using the Python application.
Data accessibility	Over-The-Counter medicine data used in this data set is collected from pharmacies: - Apotek Primedika-Jongke, Sleman, Daerah Istimewa Yogyakarta, Indonesia - Apotek Bina Farma, Sleman, Daerah Istimewa Yogyakarta, Indonesia - Apotek Sinar Farma, Wonosari, Daerah Istimewa Yogyakarta, Indonesia Books: - Informasi Socialite Obat Indonesia Repository name: Mendeley Data: Pharma-safe Index Dataset [1] Data identification number: 10.17632/m78j7ctwhn.1 Direct URL to data: https://data.mendeley.com/datasets/m78j7ctwhn/2
Related research article	

1. Value of the Data

- The data can be used in a knowledge-based system for the public, providing easy access to accurate and comprehensive information about medications, including proper dosages and potential side effects. It educates the public about contraindications, helping them avoid harmful drug interactions and allergic reactions. This empowers individuals to make informed decisions about their healthcare by understanding the benefits and risks associated with their medications. By offering detailed guidance, the system promotes medication safety and adherence, ultimately enhancing public health and enabling better self-management of chronic conditions.
- Currently in Indonesia there is no database that provides complete information about over-the-counter medicines, most of which only contain the active ingredients [2]. This rich dataset is crucial for training AI dan machine learning models that can optimize treatment plans and improve diagnostic tools especially for diseases that can be treated with over-the-counter medicines, especially diseases that often occur in tropical areas like Indonesia [3].
- These data can be reused by researchers as a rich source of real-world information on drug efficacy, safety, and patient outcomes, especially in Indonesia. Researchers can analyze these data to identify trends in medication use and emerging resistance patterns, providing insights that guide future research directions. This can lead to improved drug formulations, tailored treatment plans, and better public health policies. Additionally, the data can support studies on the effectiveness of different medications in diverse populations, contributing to a deeper understanding of how various factors impact drug performance and patient health.

2. Background

The rise in the utilization of over-the-counter medicine and the restricted availability of over-the-counter medicine for self-treatment can be attributed to various factors. These factors include the widespread access to information on the internet, the expensive nature of medical

examinations at healthcare facilities, and the challenges associated with accessing healthcare services, such as distance and limited service hours [4,5]. There are several diseases that can be treated independently, such as flu, coughs, fever, pain, diarrhea, worm infections and gastritis. Over-the-counter medicines are accessible without the need for examination by a doctor or the use of a prescription, so there is a risk of errors in their use [6,7].

Accurate data sets are required in the fields of pharmacology, public health, and information science to validate the efficacy of over-the-counter drugs for self-medication [8]. From a pharmacological perspective, it is essential to comprehend the active components, modes of operation, and possible interactions of over-the-counter medications in order to guarantee their safe utilization and reduce the likelihood of adverse reactions. In public health theory emphasize the importance of informed self-care and the prevention of medication misuse and adverse reactions. Information science principles guide the organization, retrieval, and dissemination of data to maximize accessibility and usability.

3. Data Description

This “Pharma-safe Index Dataset” [1] is an over-the-counter drugs dataset consisting of 112 drugs data on medicinal products sold in the three pharmacist in Yogyakarta Special Region, Indonesia. The Pharma-safe Index dataset is available in Indonesian and English. The dataset is available in 2 formats, namely JSON and CSV formats. The two file formats are separated into 2 folders in the dataset (Table 1).

Table 1
Folder structure.

Folder Structure	
indonesian	
├	csv
│	├ data-rule.csv
│	├ data-dose.csv
│	├ data-sideeffect.csv
│	├ data-price.csv
│	├ data-druginteraction.csv
│	├ data-usedperiode.csv
│	├ data-composition.csv
│	├ data-contraindication.csv
│	└ data-drug.csv
└	json
└	└ data-drugs.json
english	
├	csv
│	├ data-aturan.csv
│	├ data-dosis.csv
│	├ data-efeksamping.csv
│	├ data-harga.csv
│	├ data-interaksiobat.csv
│	├ data-jangkawaktupenggunaan.csv
│	├ data-komposisi.csv
│	├ data-kontraindikasi.csv
│	└ data-obat.csv
└	json
└	└ data-obat.json

Table 2 shows the JSON Structure of the dataset. There is no difference in the JSON structure for Indonesian and English, what's different is the key-value pair data in it. There are 11 name-value pairs in JSON level 1, namely *drug_name*, *disease_category*, *price*, *indication*, *composition*, *side_effects*, *contra_indications*, *contra_indications*, *drug_interactions*, *dosage*, *usage_rules*, and *usage_period*. Data in the Pharma-safe Index already has a structured data form and reflects the relationships in each data record.

In the dataset there is some data that is related to one-to-many data, therefore there are several CSV files to accommodate these relationships. Each has different columns according to the information you want to store. An explanation of the columns in each CSV file can be seen in **Table 3**. The CSV file *data-obat.csv* is the main CSV file and stores OTC drug data information. The *drug_id* column is a column that shows the relationship to the CSV file *data-obat.csv* and is a column that accommodates one-to-many relationships.

4. Experimental Design, Materials and Methods

The research process was carried out in 5 main steps (**Fig. 1**). The research began with observing the locations of the three pharmacies that were the object of research, namely Primedika Jongke Pharmacy, Bina Farma Pharmacy, and Sinar Farma Pharmacy. Next, a list of drugs sold by the three pharmacies was collected. Next, the author determines the data source. Data sources were taken from interview pharmacist and doctor, the physical brochure from medicine box, and pharmacy books. The research steps are Data Collection (1), Data Structure Correction (2), Data Cleaning (3), Validated by Doctor (4), Translate from Indonesian to English (5) and Export to CSV and JSON File (6).

1. *Data Collection*: Data collection is done in 3 ways, namely interview with pharmacist and doctor, collecting physical brochures from medicine box and collecting data from several pharmacy books [2]. The data that has been collected is stored in an Excel file format for further processing.
2. *Data Structure Correction*: Text data collected from physical brochures and several pharmacy books have very diverse data structures. The second step of this research is to standardize the data structure [9]. Standardized data structures are used to determine the information to be stored into dataset.
3. *Data Cleaning*: Every drug list and text data that has been collected is cleaned from characters that can interfere with the CSV and JSON format creation process. These characters are the quotation mark character (') and the double quotation mark (") [10]. In order not to lose the meaning of the sentence by removing characters, a back-slash character (\) is added to (') and (").
4. *Validated by Doctor*: After the data structure correction and data cleaning stages, the dataset is then validated by a doctor. The doctor examines the relationship of each existing data based on existing medical understanding. If there is an incorrect data, the data will be immediately corrected by a doctor based on medical knowledge.
5. *Translate from Indonesian to English*: Lastly, we translate the dataset from Bahasa Indonesian to English. The translation process is carried out manually by researchers and doctors. Some translations are adapted to language commonly used in the medical area.
6. *Export to CSV & JSON*: After the data is built, the final process is to create a dataset file in CSV and JSON format. A small program using the Python programming language was developed to build CSV and JSON files. The JSON structure can be seen at **Table 1** and CSV file data can be seen at **Table 2**. The JSON and CSV data formats are data formats that have high access flexibility, where both formats can be used on various platforms [11].

Table 2
JSON structure.

JSON Structure (Indonesian)	JSON Structure (English)
<pre> { nama_obat: "Antasida Doen Tablet Kunyah" kategori_penyakit: "lambung" harga: [{ bentuk_dosis: "strip" harga: 2000 }, { bentuk_dosis: "box" harga: 16000 },] indikasi: ["Obat sakit maag untuk mengurangi nyeri lambung yang disebabkan oleh kelebihan asam lambung dengan gejala seperti mual, perih kembung."] komposisi: [{ nama: "Aluminium Hydroxide" jumlah: "200" satuan: "mg" }, { nama: "Magnesium Hydroxide" jumlah: "200" satuan: "mg" }] efek_samping: ["Kadang-kadang terjadi konstipasi, diare, mual, muntah", "Hypophosphatemia osteomalacia (pada pemberian dosis besar untuk jangka lama bila kadar phosphate rendah)", "Dialysis dementsis pada pemberian jangka panjang pada pasien dialisis"] kontra_indikasi: ["Disfungsi ginjal berat", "hipersensitif", "obstruksi saluran pencernaan", "pasien yang sangat lemah"] interaksi_obat: ["Aluminium hydroxide akan mengurangi absorpsi bermacam- macam vitamin Tetracyclin",] dosis: { dewasa: "1-2 tablet, 3-4 kali per hari." anak-anak: "0.5-1 tablet, 3-4 kali per hari" } aturan_pakai: ["Sebaiknya diberikan pada saat perut kosong", "Berikan 1-2 jam setelah makan sebelum tidur", "Kunyah tablet degan baik sebelum ditelan"] jangka_waktu_penggunaan: ["Jangan digunakan selama 2 minggu"] }, ... </pre>	<pre> { drug_name: "Antasida Doen Tablet Kunyah" disease_category: "stomach" price: [{ form: "strip" price: 2000 }, { form: "box" price: 16000 },] indication: ["gastric ulcer medicine to reduce gastric pain caused by excess gastric acid with symptoms such as nausea, bloating."] composition: [{ name: "Aluminium Hydroxide" amount: "200" unit: "mg" }, { name: "Magnesium Hydroxide" amount: "200" unit: "mg" }] side_effect: ["Constipation, diarrhea Nauseous, vomit may occured", "Hypophosphatemia osteomalacia (on long-term administration of large doses when phosphate levels are low)", "Dialysis dementia in long-term administration in dialysis patients"] contra_indication: ["Severe kidney dysfunction", "hypersensitive", "digestive tract obstruction", "very weak patient"] drug_interaction: [" Aluminium hydroxide will reduce the absorption of various vitamins and Tetracyclin ",] dose: { adult: "1-2 tablets, 3-4 times per day." children: "0.5-1 tablet, 3-4 times per day" } aturan_pakai: [" Best given on an empty stomach", " Give 1-2 hours after eating before bed", " Chew the tablet well before swallowing"] jangka_waktu_penggunaan: ["Do not use for 2 weeks"] }, ... </pre>

Table 3
Overview of column structure in CSV files.

No	CSV File	CSV Column	Description
1	data-obat.csv	id	id or medicine code
		nama_obat	Name of the medicine
		kategori_penyakit	Medicine category
2	data-aturan.csv	indikasi	Indiation
		id_obat	Id or medicine code; relation to data-obat.csv
3	data-dosis.csv	aturan_pakai	How to use medicine
		id_obat	Id or medicine code; relation to data-obat.csv
4	data-efeksamping.csv	profil	Profile type of medicine user
		dosis	Dose
		id_obat	Id or medicine code; relation to data-obat.csv
5	data-harga.csv	efek_samping	Side effect
		id_obat	Id or medicine code; relation to data-obat.csv
6	data-interaksiobat.csv	satuan	Medicine unit
		harga	Medicine price
		id_obat	Id or medicine code; relation to data-obat.csv
7	data-jangkawaktupenggunaan.csv	Interaksi_obat	Medicine interaction
		id_obat	Id or medicine code; relation to data-obat.csv
8	data-komposisi.csv	jangka_waktu_penggunaan	Period of use
		id_obat	Id or medicine code; relation to data-obat.csv
		nama	Compound name
		jumlah	Quantity of compound
9	data-kontraindikasi.csv	satuan	Compound unit
		id_obat	Id or medicine code; relation to data-obat.csv
		kontra_indikasi	Contra indication

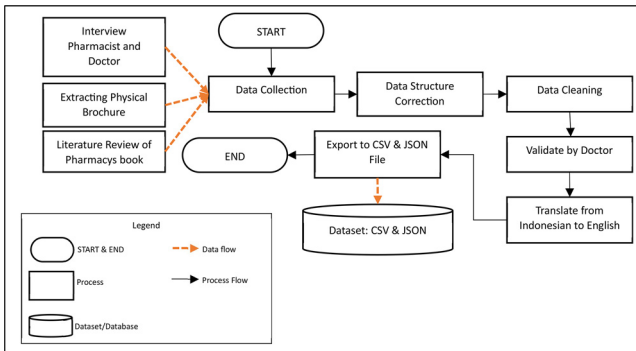


Fig. 1. Flowchart illustrating the research process.

Limitations

Currently, the dataset is limited to Over-The-Counter/OTC drugs available at pharmacies in Yogyakarta Special Region, Indonesia (Apotek Primedika Jongke, Apotek Bina Farma, and Apotek Sinar Farma). The medicines chosen are medicines used for coughs, flu and stomachache in children and adults. There are still limitations regarding the data structure in some parts, such as drug dosage and side effects. Data Structure Standardize of drug dosages is difficult because there are many combinations of age divisions (1). For example, some say the children’s ages are between 2 and 5 years, 2 and 6 years, over 4 years, and so on. Another challenge arises from the extensive usage of unstructured text in the side effects section, which often comprises complex words (2). For instance, in a certain brochure, the composition comprises the ramifications and the likelihood of its occurrence. While in other brochure, the structure is comprised of outcomes that are dependent on the dosage and usage, taking into account any pre-existing

conditions. Therefore, it is difficult to classify and distinguish different side effects into separate categories.

Ethics Statement

The authors hereby confirm that we have thoroughly read and adhere to the ethical guidelines for publishing in Data in Brief. Furthermore, we affirm that the ongoing project does not entail the use of human subjects, animal experimentation, or any data obtained from social media platforms.

CRediT Author Statement

Danny Sebastian: Conceptualization, Methodology, Software, Writing – Original Draft; **Restyandito:** Conceptualization, Methodology, Writing – Review & Editing; **Justinus Putranto Agung Nugroho:** Resources, Validation, Writing – Review & Editing.

Data Availability

[Pharma-safe Index Dataset \(Original data\)](#) (Mendeley Data).

Acknowledgments

Medicine Knowledge Contributors: Apt. Assysifa Septyani Putri, S.Farm., as Pharmacist, apt. Fretty Widadi, S.Farm., as Pharmacist, Apt. Aloysius Bimo Tiar Nugroho, S.Farm., as Pharmacist, dr. Oscar Gilang Purnajati, MHPE as Validator. This research is supported by Institute of Research and Community of Service Universitas Kristen Duta Wacana (LPPM-UKDW), Contract Number: 109/D01/LPPM/2024.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] D. Sebastian, R. Restyandito, J.P.A. Nugroho, Pharma-safe index dataset, Mendeley Data (2024), doi:[10.17632/m78j7ctwhn.2](#).
- [2] Ikatan Apoteker Indonesia, Informasi Socialite Obat Indonesia, vol. 53. PT Pharma Tekno Solusi, 2021.
- [3] P. Rajpurkar, E. Chen, O. Banerjee, E.J. Topol, AI in health and medicine, *Nat. Med.* 28 (1) (2022) 31–38, doi:[10.1038/s41591-021-01614-0](#).
- [4] S.V. Brahmabhatt, V.D. Dave, A survey study on knowledge, attitude, and practice toward self-medication practice with over-the-counter drugs among under graduate dental students, *Asian J. Pharm. Res. Health Care* 15 (4) (2023) 359–363, doi:[10.4103/ajprhc.ajprhc_79_23](#).
- [5] O.A. Mirdad, et al., Over-the-counter medication use among parents in Saudi Arabia, *Int. J. Environ. Res. Public Health* 20 (2) (2023) 1193, doi:[10.3390/ijerph20021193](#).
- [6] S. Tesfamariam, et al., Self-medication with over the counter drugs, prevalence of risky practice and its associated factors in pharmacy outlets of Asmara, Eritrea, *BMC Public Health* 19 (1) (2019) 159, doi:[10.1186/s12889-019-6470-5](#).
- [7] F. Feli, L. Pratiwi, S. Rizkifani, Analisis Tingkat Pengetahuan Mahasiswa Program Studi Farmasi Terhadap Swamedikasi Obat Bebas dan Bebas Terbatas, *J. Syifa Sci. Clin. Res.* 4 (2) (2022) 275–286, doi:[10.37311/jsscr.v4i2.14027](#).

- [8] E.E. Roughead, R. Lim, Preventing overdoses with over-the-counter medicines, *Med. J. Aust.* 218 (9) (2023) 399–400, doi:[10.5694/mja2.51927](https://doi.org/10.5694/mja2.51927).
- [9] M.K. Gupta, P. Chandra, A comprehensive survey of data mining, *Int. J. Inf. Technol.* 12 (4) (2020) 1243–1257, doi:[10.1007/s41870-020-00427-7](https://doi.org/10.1007/s41870-020-00427-7).
- [10] J. Wang, X. Wang, Y. Yang, H. Zhang, B. Fang, A review of data cleaning methods for web information system, *Comput., Mater. Continua* 62 (3) (2020) 1053–1075, doi:[10.32604/cmc.2020.08675](https://doi.org/10.32604/cmc.2020.08675).
- [11] F. Pezoa, J.L. Reutter, F. Suarez, M. Ugarte, D. Vrgoč, Foundations of JSON schema, in: *Proceedings of the 25th International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland, Apr. 2016, pp. 263–273, doi:[10.1145/2872427.2883029](https://doi.org/10.1145/2872427.2883029). International World Wide Web Conferences Steering Committee.