

Journal Pre-proof

Structured Over-The-Counter Medicine Dataset to Improve the safety of Pharmaceuticals

Danny Sebastian Conceptualization , Restyandito Conceptualization , Justinus Putranto Agung Nugroho

PII: S2352-3409(24)00757-1
DOI: <https://doi.org/10.1016/j.dib.2024.110794>
Reference: DIB 110794



To appear in: *Data in Brief*

Received date: 16 July 2024
Revised date: 19 July 2024
Accepted date: 29 July 2024

Please cite this article as: Danny Sebastian Conceptualization , Restyandito Conceptualization , Justinus Putranto Agung Nugroho , Structured Over-The-Counter Medicine Dataset to Improve the safety of Pharmaceuticals, *Data in Brief* (2024), doi: <https://doi.org/10.1016/j.dib.2024.110794>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Inc.

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

1 ARTICLE INFORMATION

2 Article title

3 Structured Over-The-Counter Medicine Dataset to Improve the safety of Pharmaceuticals.

4 Authors

5 Danny – Sebastian^{a*}

6 Restyandito^a

7 Justinus Putranto Agung – Nugroho^b

8 Affiliations

9 ^aFakultas Teknologi Informasi, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia

10 ^bFakultas Kedokteran, Universitas Kristen Duta Wacana, Yogyakarta, Indonesia

11 Corresponding author's email address and Twitter handle

12 danny.sebastian@staff.ukdw.ac.id

13 Keywords

14 over-the-counter medicine dataset; self-medication; Indonesian medicine dataset; decision support
15 system; expert system

16 Abstract

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

29 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 Raw

Data collection	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 source location	Over-The-Counter medicine data used in this data set is collected from pharmacies: <ul style="list-style-type: none"> - 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: <ul style="list-style-type: none"> - Informasi Socialite Obat Indonesia
Data accessibility	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	

30

31

32 VALUE OF THE DATA

- 33 • The data can be used in a knowledge-based system for the public, providing easy
34 access to accurate and comprehensive information about medications, including
35 proper dosages and potential side effects. It educates the public about
36 contraindications, helping them avoid harmful drug interactions and allergic
37 reactions. This empowers individuals to make informed decisions about their
38 healthcare by understanding the benefits and risks associated with their medications.
39 By offering detailed guidance, the system promotes medication safety and
40 adherence, ultimately enhancing public health and enabling better self-management
41 of chronic conditions.
- 42 • Currently in Indonesia there is no database that provides complete information about
43 over-the-counter medicines, most of which only contain the active ingredients[2].

44 This rich dataset is crucial for training AI dan machine learning models that can
 45 optimize treatment plans and improve diagnostic tools especially for diseases that
 46 can be treated with over-the-counter medicines, especially diseases that often occur
 47 in tropical areas like Indonesia[3].

- 48 • These data can be reused by researchers as a rich source of real-world information on
 49 drug efficacy, safety, and patient outcomes, especially in Indonesia. Researchers can
 50 analyze these data to identify trends in medication use and emerging resistance
 51 patterns, providing insights that guide future research directions. This can lead to
 52 improved drug formulations, tailored treatment plans, and better public health
 53 policies. Additionally, the data can support studies on the effectiveness of different
 54 medications in diverse populations, contributing to a deeper understanding of how
 55 various factors impact drug performance and patient health.

56

57

58 BACKGROUND

59 The rise in the utilization of over-the-counter medicine and the restricted availability of over-the-
 60 counter medicine for self-treatment can be attributed to various factors. These factors include the
 61 widespread access to information on the internet, the expensive nature of medical examinations at
 62 healthcare facilities, and the challenges associated with accessing healthcare services, such as
 63 distance and limited service hours [4], [5]. There are several diseases that can be treated
 64 independently, such as flu, coughs, fever, pain, diarrhea, worm infections and gastritis. Over-the-
 65 counter medicines are accessible without the need for examination by a doctor or the use of a
 66 prescription, so there is a risk of errors in their use[6], [7].

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

75

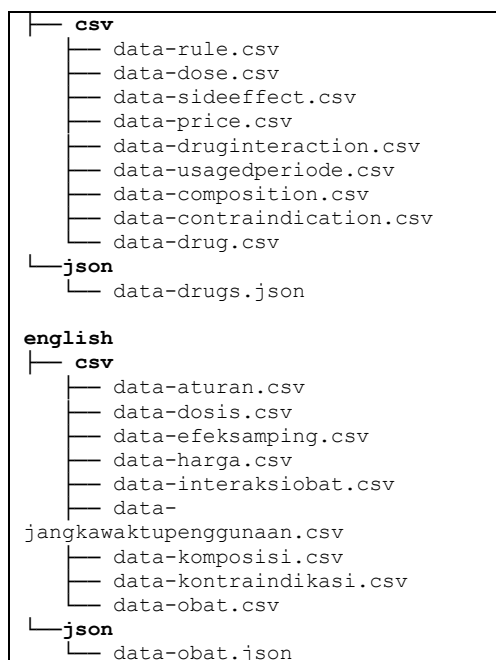
76 DATA DESCRIPTION

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

82 **Table 1**

83 Folder structure

Folder Structure
indonesian



84

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

91 **Table 1**
 92 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" │ │ } │ └──] }] </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" │ │ } │ └──] }] </pre>

```

]
└─ efek_samping: [
└─ "Kadang-kadang terjadi konstipasi,
diare, mual, muntah",
└─ "Hypophosphatemia osteomalacia (pada
pemberian dosis besar untuk jangka lama bila
kadar phosphate rendah)",
└─ "Dialysis dementia 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"
]
},
...
]
└─ 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: [
└─ "Aluminum 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"
]
},
...
]

```

93

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

100

Table 3

101

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
		indikasi	Indiation
2	data-aturan.csv	id_obat	Id or medicine code; relation to data-obat.csv
		aturan_pakai	How to use medicine
3	data-dosis.csv	id_obat	Id or medicine code; relation to data-obat.csv
		profil	Profile type of medicine user
		dosis	Dose
4	data-efeksamping.csv	id_obat	Id or medicine code; relation to data-obat.csv
		efek_samping	Side effect
5	data-harga.csv	id_obat	Id or medicine code; relation to data-obat.csv
		satuan	Medicine unit
		harga	Medicine price
6	data-interaksiobat.csv	id_obat	Id or medicine code; relation to data-obat.csv

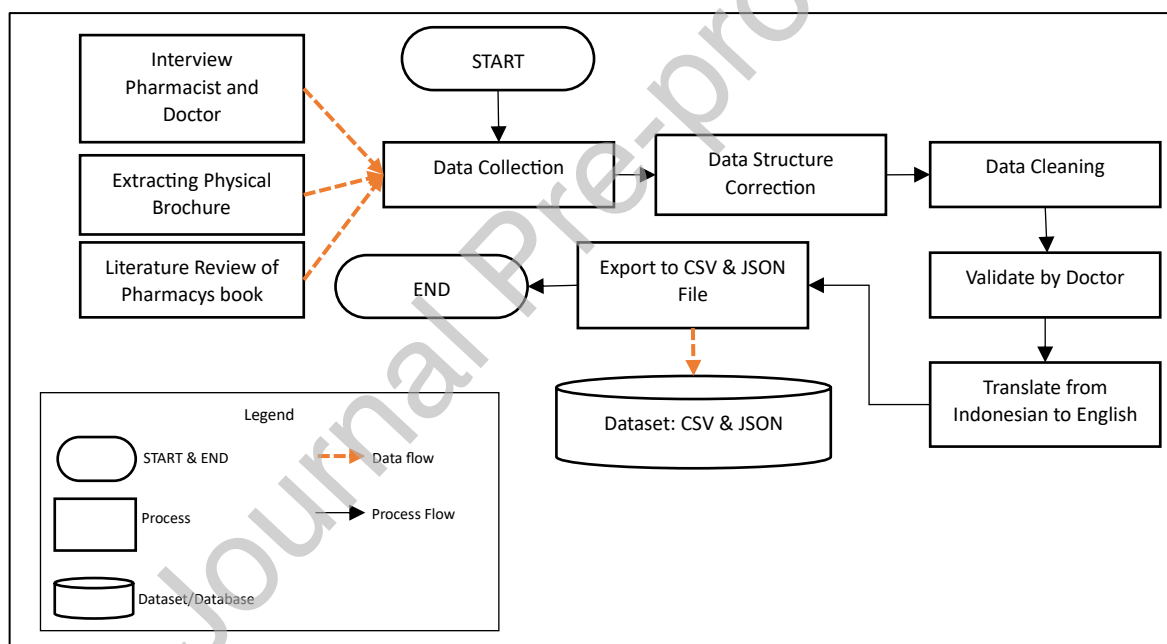
		Interaksi_obat	Medicine interaction
7	data-jangkawaktupenggunaan.csv	id_obat	Id or medicine code; relation to data-obat.csv
		jangka_waktu_penggunaan	Period of use
8	data-komposisi.csv	id_obat	Id or medicine code; relation to data-obat.csv
		nama	Compound name
		jumlah	Quantity of compound
		satuan	Compound unit
9	data-kontraindikasi.csv	id_obat	Id or medicine code; relation to data-obat.csv
		kontra_indikasi	Contra indication

102

103

104 EXPERIMENTAL DESIGN, MATERIALS AND METHODS

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



112

113 **Fig.1.** Flowchart illustrating the research process

114 **1. Data Collection:** Data collection is done in 3 ways, namely interview with pharmacist and doctor,
 115 collecting physical brochures from medicine box and collecting data from several pharmacy books[2].
 116 The data that has been collected is stored in an Excel file format for further processing.

117 **2. Data Structure Correction:** Text data collected from physical brochures and several pharmacy
 118 books have very diverse data structures. The second step of this research is to standardize the data
 119 structure[9]. Standardized data structures are used to determine the information to be stored into
 120 dataset.

121 **3. Data Cleaning:** Every drug list and text data that has been collected is cleaned from characters that
 122 can interfere with the CSV and JSON format creation process. These characters are the quotation

123 mark character (') and the double quotation mark (")[10]. In order not to lose the meaning of the
124 sentence by removing characters, a back-slash character (\) is added to (\') and (\").

125 4. **Validated by Doctor:** After the data structure correction and data cleaning stages, the dataset is
126 then validated by a doctor. The doctor examines the relationship of each existing data based on
127 existing medical understanding. If there is an incorrect data, the data will be immediately corrected
128 by a doctor based on medical knowledge.

129 5. **Translate from Indonesian to English:** Lastly, we translate the dataset from Bahasa Indonesian to
130 English. The translation process is carried out manually by researchers and doctors. Some
131 translations are adapted to language commonly used in the medical area.

132 6. **Export to CSV & JSON:** After the data is built, the final process is to create a dataset file in CSV and
133 JSON format. A small program using the Python programming language was developed to build CSV
134 and JSON files. The JSON structure can be seen at Table 1 and CSV file data can be seen at Table 2.
135 The JSON and CSV data formats are data formats that have high access flexibility, where both formats
136 can be used on various platforms[11].

137

138 LIMITATIONS

139 Currently, the dataset is limited to Over-The-Counter/OTC drugs available at pharmacies in
140 Yogyakarta Special Region, Indonesia (Apotek Primedika Jongke, Apotek Bina Farma, and Apotek
141 Sinar Farma). The medicines chosen are medicines used for coughs, flu and stomachache in children
142 and adults. There are still limitations regarding the data structure in some parts, such as drug dosage
143 and side effects. Data Structure Standardize of drug dosages is difficult because there are many
144 combinations of age divisions (1). For example, some say the children's ages are between 2-5 years,
145 2-6 years, over 4 years, and so on. Another challenge arises from the extensive usage of unstructured
146 text in the side effects section, which often comprises complex words (2). For instance, in a certain
147 brochure, the composition comprises the ramifications and the likelihood of its occurrence. While in
148 other brochure, the structure is comprised of outcomes that are dependent on the dosage and
149 usage, taking into account any pre-existing conditions. Therefore, it is difficult to classify and
150 distinguish different side effects into separate categories.

151

152 ETHICS STATEMENT

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

156

157 CRediT AUTHOR STATEMENT

158 **Danny Sebastian:** Conceptualization, Methodology, Software, Writing – Original Draft

159 **Restyandito:** Conceptualization, Methodology, Writing – Review & Editing

160 **Justinus Putranto Agung Nugroho:** Resources, Validation, Writing – Review & Editing

161

162 **ACKNOWLEDGEMENTS**

163 **Medicine Knowledge Contributors:** Apt. Assysifa Septyani Putri, S.Farm., as Pharmacist, apt. Fretty
164 Widadi, S.Farm., as Pharmacist, Apt. Aloysius Bimo Tiar Nugroho, S.Farm., as Pharmacist, dr. Oscar
165 Gilang Purnajati, MHPE as Validator.

166 This research is supported by Institute of Research and Community of Service Universitas Kristen
167 Duta Wacana (LPPM-UKDW), Contract Number: 109/D01/LPPM/2024.

168

169

170 **DECLARATION OF COMPETING INTERESTS**

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

173

174

175 **REFERENCES**

- 176 [1] D. Sebastian, R. Restyandito, and J. P. A. Nugroho, "Pharma-safe Index Dataset." Mendeley
177 Data, 2024. doi: 10.17632/m78j7ctwhn.2.
- 178 [2] Ikatan Apoteker Indonesia, Informasi Socialite Obat Indonesia, vol. 53. PT Pharma Tekno Solusi,
179 2021.
- 180 [3] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nat Med*, vol.
181 28, no. 1, pp. 31–38, Jan. 2022, doi: 10.1038/s41591-021-01614-0.
- 182 [4] S. V. Brahmhatt and V. D. Dave, "A Survey Study on Knowledge, Attitude, and Practice toward
183 Self-medication Practice with Over-the-counter Drugs among Under Graduate Dental
184 Students," *Asian Journal of Pharmaceutical Research and Health Care*, vol. 15, no. 4, pp. 359–
185 363, 2023, doi: 10.4103/ajprhc.ajprhc_79_23.
- 186 [5] O. A. Mirdad et al., "Over-the-Counter Medication Use among Parents in Saudi Arabia," *Int J*
187 *Environ Res Public Health*, vol. 20, no. 2, p. 1193, Jan. 2023, doi: 10.3390/ijerph20021193.
- 188 [6] S. Tesfamariam et al., "Self-medication with over the counter drugs, prevalence of risky
189 practice and its associated factors in pharmacy outlets of Asmara, Eritrea," *BMC Public Health*,
190 vol. 19, no. 1, p. 159, Dec. 2019, doi: 10.1186/s12889-019-6470-5.
- 191 [7] F. Feli, L. Pratiwi, and S. Rizkifani, "Analisis Tingkat Pengetahuan Mahasiswa Program Studi
192 Farmasi Terhadap Swamedikasi Obat Bebas dan Bebas Terbatas," *Journal Syifa Sciences and*
193 *Clinical Research*, vol. 4, no. 2, pp. 275–286, Apr. 2022, doi: 10.37311/jsscr.v4i2.14027.
- 194 [8] E. E. Roughead and R. Lim, "Preventing overdoses with over-the-counter medicines," *Medical*
195 *Journal of Australia*, vol. 218, no. 9, pp. 399–400, May 2023, doi: 10.5694/mja2.51927.
- 196 [9] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *International Journal of*
197 *Information Technology*, vol. 12, no. 4, pp. 1243–1257, Dec. 2020, doi: 10.1007/s41870-020-
198 00427-7.

- 199 [10] J. Wang, X. Wang, Y. Yang, H. Zhang, and B. Fang, "A Review of Data Cleaning Methods for Web
200 Information System," *Computers, Materials & Continua*, vol. 62, no. 3, pp. 1053–1075, 2020,
201 doi: 10.32604/cmc.2020.08675.
- 202 [11] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, and D. Vrgoč, "Foundations of JSON Schema," in
203 Proceedings of the 25th International Conference on World Wide Web, Republic and Canton of
204 Geneva, Switzerland: International World Wide Web Conferences Steering Committee, Apr.
205 2016, pp. 263–273. doi: 10.1145/2872427.2883029.

Journal Pre-proof