An Indonesian Physical Commonsense Reasoning Dataset (PIQA-style)

Muhammad Ravi Shulthan Habibi
Universitas Indonesia
muhammadravi251001@gmail.com

Abstract

We present a physical commonsense reasoning dataset for Indonesian, constructed in the PIQA-style format. The dataset contains 250 original examples designed to evaluate model performance in understanding everyday physical knowledge. The dataset aims to provide a valuable benchmark for multilingual NLP research, particularly in under-resourced languages.

1 Introduction

Many non-English languages lack culturally specific evaluation datasets created by native speakers. Physical commonsense reasoning is especially important, as it tests whether models can apply world knowledge to everyday tasks, such as cooking, cleaning, and basic physical processes.

We introduce an Indonesian dataset of 250 PIQA-style items. Each item consists of a prompt, two candidate solutions, and a label indicating the correct answer. The dataset contributes to the broader effort of building multilingual benchmarks for commonsense reasoning.

2 Related Work

The most relevant prior work is the **PIQA dataset** (Bisk et al., 2020), which introduced a benchmark for physical commonsense reasoning in English. PIQA consists of prompts paired with two possible solutions, with the task of choosing the physically plausible one. Our dataset follows the same format but is newly created for Indonesian, ensuring cultural and linguistic relevance. Other commonsense benchmarks such as **CommonsenseQA** (Talmor et al., 2019) focus on general knowledge, but few exist for physical reasoning in non-English languages.

3 Dataset Construction

Format: Each example contains a prompt, two solutions (`solution0`, `solution1`), and a label (0 or 1). **Size:** 250 examples. **Method:** Examples were generated with the assistance of a large language model (ChatGPT, `gpt-5`) and carefully guided by prompts to produce PIQA-style items. All items were manually reviewed, corrected, and finalized by a native speaker of Indonesian to ensure quality, correctness, and cultural relevance. In particular, at least 50 examples reflect uniquely Indonesian contexts (e.g., local foods, household practices, and traditional objects). **Label Distribution:** 125 with label 0, and 125 with label 1.

Label	Count
0	125
1	125

Table 1: Label distribution in the dataset.

4 Dialect Information

The dataset uses Standard Indonesian (Bahasa Indonesia), without influence from regional dialects.

5 Examples

Table 2 shows ten examples from the dataset.

6 Verification

All examples were verified by at least one native speaker of Indonesian (the author).

7 Conclusion

We present a new dataset for physical commonsense reasoning in Indonesian. We hope this resource will serve as a useful benchmark for future research on multilingual commonsense reasoning.

Prompt	Solution0	Solution1	Label
Bagaimana cara memadamkan lilin yang menyala dengan aman?	Tiup lilin sampai api padam.	Siram lilin dengan air sampai api padam.	0
Jika seseorang berhenti makan dalam waktu lama, apa akibatnya?	Tubuh semakin sehat.	Tubuh menjadi lemas.	1
Jika motor dipacu terlalu cepat di jalan licin, apa akibatnya?	Bisa tergelincir.	Motor jadi kebal air.	1
Bagaimana cara membuat lampu menyala?	Colokkan ke listrik dan tekan tombol on.	Letakkan lampu di atas meja tanpa kabel.	0
Apa cara yang benar untuk menggunakan gunting?	Pegang gagang dan tekan bilah untuk memotong.	Tarik bilah dengan tangan kosong.	0
Jika nasi uduk disimpan semalaman di suhu ruang, apa risikonya?	Nasi bisa basi dan berbau asam.	Nasi jadi lebih segar.	0
Mengapa bambu berongga sering digunakan untuk membuat kentongan?	Karena menghasilkan suara nyaring saat dipukul.	Karena bisa mengeluarkan air saat dipukul.	0
Untuk membuat tape singkong berhasil, apa yang harus ditambahkan?	Ragi tape.	Garam dalam jumlah banyak.	0
Mengapa orang Jawa menaburkan garam di sekitar rumah saat ada ulat grayak menyerang sawah?	Agar tanah jadi lebih subur.	Untuk mencegah hama merayap masuk.	1
Apa fungsi kipas tangan dari anyaman bambu saat mengipas sate?	Membuat arang padam.	Membuat arang tetap menyala.	1
Jika batok kelapa dibakar, apa yang terjadi?	Menghasilkan air segar.	Menghasilkan bara panas dan asap.	1

Table 2: Ten example items from the Indonesian PIQA-style dataset.

References

Yonatan Bisk, Rowan Zellers, Jianfeng Gao, and Yejin Choi. 2020. Piqa: Reasoning about physical commonsense in natural language. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. 2019. Commonsenseqa: A question answering challenge targeting commonsense knowledge. In *Proceedings of NAACL-HLT*.