

# Physical Commonsense Reasoning Dataset in Banglish

**Author:** Rishav Ghosh

**Contact:** rishavghosh605@gmail.com

## Dataset Description

This dataset contains 100 physical commonsense reasoning examples written in Banglish - Bengali language written in Latin script as commonly used by Bengali speakers online and in informal communication. The dataset covers Bengali cultural contexts including festivals like Durga Puja and Kali Puja, traditional foods like roshogolla and kochuri, and everyday objects like matir hari (clay pots) and moshari (mosquito nets).

All examples follow the PIQA format with a prompt and two solutions where only one is physically correct. For example: "Durga pujo shesh howar por matir murti ta jole bashiye dile ki hobe?" (What happens when the clay Durga idol is immersed in water after the festival?) with solutions about the clay dissolving vs. becoming harder.

## Creation Process

I created all examples manually by first thinking about different aspects of Bengali culture - festivals, cooking, traditional games, household items, and seasonal practices. For each cultural domain, I thought of scenarios that involved physical reasoning about things like material properties, temperature, weight, or how objects behave when they interact.

I used ChatGPT to help brainstorm cultural topics I might have missed and to double-check whether my examples actually involved physical reasoning rather than just cultural knowledge. But I didn't copy or generate any examples directly from ChatGPT - everything was written by hand to make sure the Banglish sounded natural and the cultural details were accurate.

The examples cover different types of physical reasoning: what happens when clay meets water, how hot oil affects food, why metal is stronger than wood, how wind affects flames, and similar everyday physics that any Bengali speaker would know from experience.

## Language and Cultural Details

Banglish preserves Bengali grammar and sentence structure while using Latin script. Words are written phonetically (like "roshogolla" for রসগোল্লা) and the examples include culture-specific concepts that wouldn't translate directly to English, like dhunuchi dancing with burning incense or alpona paste made from rice powder.

The examples assume knowledge that's common among Bengali speakers - understanding festivals, traditional cooking methods, regional games, and household practices. This makes the dataset culturally specific rather than just translated versions of universal concepts.

## Dataset Statistics

The 100 examples cover major Bengali cultural domains:

- Religious festivals and practices (Durga Puja, Kali Puja, Diwali): 25 examples
- Traditional foods and cooking: 20 examples
- Household objects and tools: 18 examples
- Traditional games and activities: 15 examples
- Seasonal practices and nature: 12 examples
- Folk traditions and customs: 10 examples

Most examples are over 25 words long to avoid being too simple. The two solution choices typically differ by just one or two key words to make the reasoning challenging but clear. Labels are balanced with 50 examples each for correct solutions being option 0 or 1.

## Reproduction Notes

To create a similar dataset, someone would need native Bengali language skills and familiarity with Bengali culture. The process involves identifying cultural practices that naturally involve physical reasoning, then writing scenarios in natural Banglish that test understanding of basic physics within those cultural contexts.

The key is making sure the reasoning depends on physical properties (like clay dissolving in water) rather than memorizing cultural facts (like which day a festival occurs). A Bengali speaker could recreate this by systematically going through cultural domains and thinking of physical interactions within each one.

---

*Dataset file: MRL Bengali - Dataset.tsv (100 examples in TSV format)*