

Myanmar–Agriculture–1K

Min Si Thu, Khin Myat Noe

August 2024

1 Introduction

Myanmar is an agricultural country, that mainly grows rice all over the country. But there is no single open source dataset about Agriculture Question–Answering in the Burmese Language as far as my knowledge reaches.

2 Purpose

The "Myanmar Agriculture 1K" Dataset is curated to build a knowledge bank for further studies in Natural Language Processing in the Burmese Language and to train instruction fine–tuned language model for the Burmese Language.

Moreover, this dataset is a motivation to move the Burmese language from a low–resource language to a high–resource language.

3 Dataset Curation Process

The dataset is curated in 2024, August by two contributors – Min Si Thu and Khin Myat Noe. The dataset contains two columns – **Instruction** and **Output**. The dataset rows are stored in **Agriculture.csv** file.

The dataset contains information about major crops and trees in Myanmar, methods of agriculture, and horticulture, chemicals and pesticides, phytopathology, climate change, and organic manure. 1053 rows of instruction, output pairs are included in the dataset.

The process of curation of the dataset has two phases. The first phase is annotating the information from the Wikipedia website "<https://www.doa.gov.mm/fes/index.php?route=common/home>". The second phase is reviewing the spelling and vocabulary mistakes in the dataset.

3.1 Examples

1. စပါးကြဲပက်စိုက်ပျိုးနည်းကို ဘယ်အခါတွေမှာ အသုံးပြုသင့်သလဲ?၊ မိုးဦးကာလနှင့် မိုးလယ်ကာလ၊ မိုးခေါင်သောအခါ၊ ရေကြီးနစ်မြုပ်ဖြစ်တတ်သော ရေနက်ကွင်းများ၊ ဒီရေအတက်အကျရှိသော နေရာများ၊ သွင်းရေရရှိရန် မလွယ်ကူသော မိုးရေသောက် ကုန်းမြင့်ပိုင်းလယ်များ၊ အချိန်လုံခြုံ စိုက်ပျိုးရန်လိုအပ်သည့် ကာလများ
2. Paris Climate Agreement ကို ဘယ်အချိန်က စတင်ချုပ်ဆိုခဲ့တာလဲ?၊ ၂၀၁၅ ခုနှစ် ဒီဇင်ဘာလမှာ ပြင်သစ်နိုင်ငံ ပါရီမြို့တွင် ကျင်းပတဲ့ COP21 ညီလာခံမှာ စတင်ချုပ်ဆိုခဲ့ပါတယ်။
3. စပါးဂုတ်ကျိုးရောဂါ၏ လက္ခဏာရပ်များကို ဖော်ပြပါ၊ ပျိုးခင်းကတည်းက ရောဂါစတင်ဖြစ်ပေါ်နိုင်သည်။ စပါးရွက်ပေါ်တွင် လွှန်းပုံသဏ္ဌာန်ရှိသော အကွက်များတွေ့ရသည်။ အကွက်များကြီးခဲ့လျှင် အလယ်တွင် မီးခိုးရောင်ပေါ်တတ်သည်။ ရောဂါကွက်တစ်ခုနှင့်တစ်ခုဆက်သွားပြီး စပါးရွက်များ သေသွားတတ်သည်။
4. ရွက်ဖြန်းအားဆေးများ၏ ဖြစ်နိုင်ချေရှိသော ဘေးထွက်ဆိုးကျိုးများကို ဖော်ပြပါ။ ရွက်ဖြန်းအားဆေးများကို မှားယွင်းစွာ အသုံးပြုပါက အပင်များ အရွက်လောင်ကျွမ်းခြင်း (leaf burn)၊ အဆိပ်သင့်ခြင်း (toxicity)၊ ပတ်ဝန်းကျင်ညစ်ညမ်းမှု (environmental pollution) နှင့် ရေအရင်းအမြစ်များ ညစ်ညမ်းခြင်း (water contamination) တို့ ဖြစ်ပွားနိုင်ပါသည်။

4 Where to download the dataset

Github - <https://github.com/MinSiThu/Myanmar-Agriculture-1K>

Zenodo - <https://zenodo.org/records/13318601>

Huggingface - <https://github.com/MinSiThu/Myanmar-Agriculture-1K>

Kaggle – <https://www.kaggle.com/datasets/minsithu/myanmar-agriculture-1k>

4.1 Applications

Myanmar Agriculture 1K Dataset can be used to build various agriculture-related NLP applications.

1. The dataset can be used for pretraining or finetuning the dataset on Burmese Large Language Models.
2. The dataset is ready to use in building RAG-based Applications.

5 Acknowledgements

Special thanks to magickospace.org for supporting the curation process of the Myanmar Agriculture 1K Dataset.

6 References for this dataset

1. <https://my.wikipedia.org/wiki/> – For Agricultural facts
2. <https://www.doa.gov.mm/fes/index.php?route=common/home> – For Agricultural facts