

InClass Prediction Competition

[Student] Shopee Code League - Product Translation

Shopee Product Title translation

2 teams · a day ago

Overview

Data

Notebooks

Discussion

Leaderboard

Rules

Team

My Submissions

Late Submission

Data Description

In Shopee Product Title Translation competition, there are two large scale product title data in traditional Chinese and English. They are from popular product categories, e.g., women's fashion, electronics, etc. The two dataset are not in labelled pair format and contain a lot of noises.

File descriptions

- train_tcn.csv: training dataset for traditional Chinese product titles (500k)
- train_en.csv: training dataset for English product titles (500k)
- dev_tcn.csv: dev (validation) set for traditional Chinese product titles (1k)

Note:

train_tcn.csv and train_en.csv are not parallel data

dev_tcn.csv and dev_en.csv are parallel data which you can check your model's sacrebleu score with lowercase parameter (<https://github.com/mjpost/sacrebleu>).

Submission file should follow the dev_en.csv format with a header of "translation_output" and all translations followed in each row.

Columns of data fields

- product_title: product title (str).
- category: product category (str).

```
> kaggle competitions download -c student-shopee-code-league-product-translation
```

Data Explorer

70.66 MB

dev_tcn.csv

test_tcn.csv

train_en.csv

train_tcn.csv

Summary

5 files

9 columns

Download All

dev_en.csv (55.19 KB)

Oppo A75 A75S A73 Phone Case Soft Rabbit Silicone Case
SOFT 99 Coating Car Wax Strong Water Watt
Low Sugar Mango Dry 250g Be The Royal
* the culture Japan Imported Round Top Space Craft - Diamond SC - MK - 010
Hello Kitty Sandals Shoes White/Red Children no739
Thunderbird Lt - 234 Japanese Wood Grain Table Mat 45 X 60 Cm / Pc
The sand Kid DVD Vol-08
Anacomda Python T1 Titan Series 120GB SSD Drive