Competition

Datathon @ IndoML 2024

Organized by shubhadip_nag - Current server time: Sept. 3, 2024, 4:15 p.m. UTC

▶ Current	Next	End
Phase-1 (Development Phase)	Phase-2 (Final Phase)	Competition Ends
Aug. 5, 2024, 7 a.m. UTC	Sept. 15, 2024, 7 a.m. UTC	Sept. 5, 2024, 7 a.m. UTC

Learn the Details | Phases | Participate | Results

Forums (/forums/19851/)

- **Overview**
- Evaluation
- Terms and Conditions

Task: Attribute-Value Prediction From E-Commerce Product Descriptions

In recent years, e-commerce has seen tremendous growth, with major online retailers offering billions of products and shipping millions of packages daily. However, given the sheer volume of offerings, sellers find it extremely difficult

to fill in extensive sets of product attributes, resulting in incomplete product profiles. E-commerce platforms, on the other hand, depend on such structured metadata, typically in the form of attribute-value pairs, for a deeper understanding of the products, and for facilitating critical downstream applications, such as search, product recommendation, question answering; as well as, for providing an enhanced customer experience.

Predicting attribute-value pairs from unstructured product descriptions is, therefore, a fundamental challenge for worldwide e-commerce catalogs such as Amazon, Walmart, and Alibaba. In this challenge, your task would be to develop a model that would automatically predict attribute-value pairs for a given product description. Along with a short product title, you will be provided details about the store and manufacturer, which you may choose not to use. You will then be required to predict the values for 5 levels of categories (starting from L0 to L4) and the brand for the given product as detailed below. Please note that the values may or may not appear in the product title. Also, unseen values for the brand/categories might appear in the hidden test data.

Datathon @ IndoML 2024

TLDR: Important dates and steps

- 🗓️ Phase-1 Deadline: 5th September 2024

12.00 Noon IST

- 🗓️ Starting Date: Phase-1 -> 5th August // Phase-2 -> 15th September
- 👉 Register: https://forms.gle/cVzA1cwyzuvXtzh17 (https://forms.gle/cVzA1cwyzuvXtzh17)
- 🚢 Join Discord Channel: https://discord.com/invite/V2W7gY8DRa (https://discord.com/invite/V2W7gY8DRa)
- 📂 Download the Data: https://codalab.lisn.upsaclay.fr/competitions/19907 (../19907) -> Participate -> Files -> Public Data
- 🏟️ Submit your predictions at: https://codalab.lisn.upsaclay.fr/competitions/19907 (../19907)
- 🚀 Starting Kit (Classification-Based): https://tinyurl.com/ymem4sxk (https://tinyurl.com/ymem4sxk)
- 🚀Starting Kit (Generative-Based) & Submission Guidelines: https://youtu.be/V5-KZNYaAEY?

si=kkXmHIOeFyeSWMTJ (https://youtu.be/V5-KZNYaAEY?si=kkXmHIOeFyeSWMTJ)

- 🚀First AMA Session: https://youtu.be/BWrDWUpj8j8 (https://youtu.be/BWrDWUpj8j8)
- 🗓️ Competition Deadline: Phase-1 -> 5th September // Phase-2 -> 22nd October
- 🏆 Win prizes worth INR 1.5 Lakhs

Phase-1 / Development Phase

Data Format

The data will be distributed in JSONL format with one example per line (see http://jsonlines.org (http://jsonlines.org/) for more details).

Each line will have an ID associated with it which will be useful for submitting predictions.

The data files have 2 types of extensions. One is '*.data' and the other one is '*.solution'

'*.data' files contain the features of that data point that will be used to train/test the model. Where, '*.solution' files contain a set of labels corresponding to that data point. Both file has a field

called "indoml_id" that will be used to map the '.data' file to the '.solution' file.

The `'.data'` file will contain the following fields:

```
indoml_id: "id_num"

id_num: id of that data point

title: Product Title

store: Store name of that product

details_Manufacturer: Manufacturer of that product
```

Example:

```
{"indoml_id": 275, "title": "Cascade Actionpacs
Dishwasher Detergent, Fresh Scent, 110 Count", "store":
"Cascade", "details_Manufacturer": "Cascade"}
```

- This example was taken from the `'attribute_train.data'` folder

The `'.solution'` file will contain the following fields:

indoml_id: "id_num" which is associated with the corresponding `'.data'` file.

```
id_num: id of that data point

details_Brand, L0_category, L1_category, L2_category, L3_category, L4_category:
These are the attribute that has to be predicted.
```

Example:

```
{"indoml_id": 275, "details_Brand": "Cascade",
"L0_category": "Health & Household", "L1_category":
"Household Supplies", "L2_category": "Dishwashing",
```

"L3_category": "Dishwasher Detergent", "L4_category": "na"}

- This example was taken from the `'attribute_train.solution'` folder

Please note that here value of "indoml_id" is an integer. Don't submit it as a string.

Answer Submission Instructions

Submit the answers file as a **`ZIP file`** containing your predictions (**`The prediction file format should be: attribute_test*.predict`**, where '*' is any string. **The best is to name your file as `'attribute_test_yourname.predict'`**. Each line in the file should be a dictionary and should adhere to the following format and the order of the `'indoml_id'` must be preserved.

There is no restriction on the zip file name. It can be named anything, such as `'submission_1.zip'`

{"indoml_id": 275, "details_Brand": "Cascade", "L0_category": "Health & Household", "L1_category": "Household Supplies", "L2_category": "Dishwashing", "L3_category": "Dishwasher Detergent", "L4_category": "na"}

Please note that here value of "indoml_id" is an integer. Don't submit it as a string.

Phase-2 / Final Phase

Oh hello!! Please finish the Phase-1 first!! 🤣

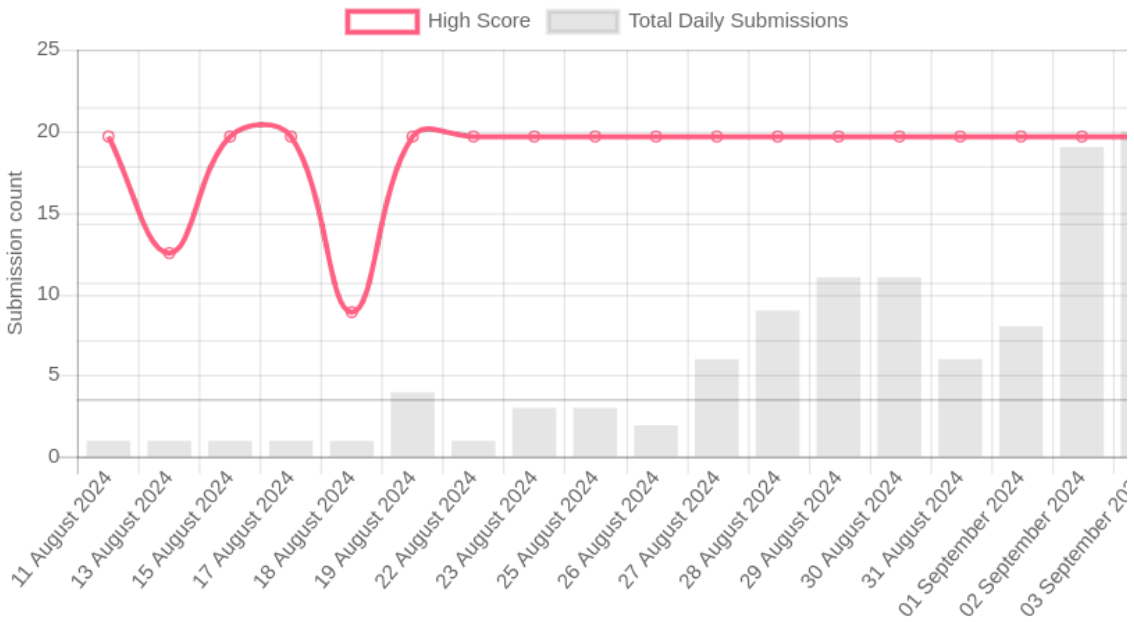
Privacy and Terms (https://github.com/codalab/codalab/wiki/Privacy) v1.6

Contact Us:

- Competition Website (https://sites.google.com/view/datathon-indoml24)
- Discord (https://discord.com/invite/V2W7gY8DRa)
- Use this link to join the Discord channel if you have not already: Discord Invite (https://discord.com/invite/V2W7gY8DRa)

(https://www.youtube.com/watch?v=VaiMflwm520&list=PLvzThk3qRkmVe1v2wRRKnQeV7BHIY5xQW&index=4)

#	Username	Score
1	neeldevenshah	0.9417
2	manoj0606	0.9413
3	vanshnawander	0.9407



Join us on Github for contact & bug reports (https://github.com/codalab/codalab/issues) About (https://github.com/codalab/codalab/wiki/Project_About_CodaLab)