00D : 01H : 12M : 37S

Data Sets and Problem Statements

Data Sets

To download the sample data input-output, Click here (https://cdn.skillenza.com/files/853304e8-7c89-4841-a7b9-ad990a721e31/sampledatainputoutput.xlsx)

To download the sample data, Click here (https://cdn.skillenza.com/files/8129c831-cb0e-493e-8d02-75e297b331bd/sampledata.xlsx)

To download the Train Data, Click here (https://cdn.skillenza.com/files/6a5fa354-63f4-4075-ae9f-ed47b60c41c7/Train-Data.xlsx)

Detailed Problem Statement:-

Developing Contextual Search for an FMCG Catalogue

Description:

You have to develop a contextual NLP based search algorithm which understands the intent behind search and points to the correct item(s) in our master catalogue. It should have the following capabilities:-

- Should work with as little as 3 characters
- Training can be onetime (time-consuming) but the response layer should return the result in <1 s
- Should be multivariate ex: Beverages with Price less than 100 Rs
- Should provide feature to boost (or give unequal) weightage to certain products/items or fields

What data will we have:

- 8k SKUs(Stock keeping unit) with
    - Product Descriptions
    - Brand Names
    - MRP
    - Promotions
    - Category Mapping

Success KPI:

The model should be able to handle queries of type

- where is and/or is ....
- There would be standard 50 such search queries and each algorithm would be measured on following
    - False Positives (Lower the better)
    - Missed Items (Lower the better)
    - Average Boosting Rank

Quantifiable indices for

- Precision=relevant results/(relevant results+ non-relevant results)

Overall = a b sqrt(c)

00D : 01H : 12M : 37S

Resubmit