Topic prediction for content-based recommender engine

**Goal of the competition**

In this competition, you will predict topics for each course based on the metadata of the course. Specifically, you will create a model that will predict up to three topics from a list of 148 topics based on the title and description of the course. A small sample file will be shared that shows some examples demonstrating the expected topics for the courses.

**Context**

In learning systems like Adobe Learning Manager, the authors create courses or import catalogs that are not properly annotated with topics/skills. The recommender system automatically maps the topics & skills to the course and utilize this information during generation of recommendation. The correct labels are often not available for training classification models, and it is usually an unsupervised learning problem. The feedback for a good topic annotation is often received from the learner’s implicit feedback during interaction with the recommendation strips.

**Dataset description**

course\_metadata.csv – Course metadata consists of title and description of the course provided by the vendors.

topics.csv - It consists of 148 topics

sample\_course\_topics\_mapping.csv – Consists of 50 examples of course and topics mapping.

**Evaluation**

F\_score\_EM – F-score is calculated based on the exact match of predicted topics against the ground truth

F\_score\_AOM – F-score is calculated based on at least one match between predicted topics and ground truth.

Evaluation metric = 0.75\*F\_score\_EM + 0.25\*F\_score\_AOM

**Submission file**

For each ‘course\_id’, you will predict up to 3 topics and add them in the ‘topics’ column, separated by “|||” delimiter.

The file contains a header row and have the following format:

course\_id,topics

12345,Business Skills|||General Management|||Business Strategy