Topic Modeling for Research Articles 2.0

Researchers have access to large online archives of scientific articles. As a consequence, finding relevant articles has become more and more difficult. Tagging or topic modelling provides a way to give clear token of identification to research articles which facilitates recommendation and search process.   
  
Earlier on the Independence Day we conducted a Hackathon to predict the topics for each article included in the test set. Continuing with the same problem, In this Live Hackathon we will take one more step ahead and predict the tags associated with the articles.  
  
Given the abstracts for a set of research articles, predict the tags for each article included in the test set.   
Note that a research article can possibly have multiple tags. The research article abstracts are sourced from the following 4 topics:

1. Computer Science

2. Mathematics

3. Physics

4. Statistics

List of possible tags are as follows:

[Tags, Analysis of PDEs, Applications, Artificial Intelligence,Astrophysics of Galaxies, Computation and Language, Computer Vision and Pattern Recognition, Cosmology and Nongalactic Astrophysics, Data Structures and Algorithms, Differential Geometry, Earth and Planetary Astrophysics, Fluid Dynamics,Information Theory, Instrumentation and Methods for Astrophysics, Machine Learning, Materials Science, Methodology, Number Theory, Optimization and Control, Representation Theory, Robotics, Social and Information Networks, Statistics Theory, Strongly Correlated Electrons, Superconductivity, Systems and Control]

**Data Dictionary**

**train.zip**

**train.csv**

|  |  |
| --- | --- |
| id | Unique ID for each article |
| ABSTRACT | Abstract of the research article |
| Computer Science | Whether article belongs to topic computer science (1/0) |
| Mathematics | Whether article belongs to topic Mathematics (1/0) |
| Physics | Whether article belongs to topic physics (1/0) |
| Statistics | Whether article belongs to topic Statistics (1/0) |
| Tags | (TARGET) There are 25 columns of possible tags with (1/0) : 1 : if article belongs to that tag 0 : if article doesn't belong to that tag |

**test.csv**

|  |  |
| --- | --- |
| id | Unique ID for each article |
| ABSTRACT | Abstract of the research article |
| Computer Science | Whether article belongs to topic computer science (1/0) |
| Mathematics | Whether article belongs to topic Mathematics (1/0) |
| Physics | Whether article belongs to topic physics (1/0) |
| Statistics | Whether article belongs to topic Statistics (1/0) |

**sample\_submission.csv**

|  |  |
| --- | --- |
| id | Unique ID for each article |
| Tags | ( Predictions) There are 25 columns of possible tags with (1/0) : 1 : if article belongs to that tag 0 : if article doesn't belong to that tag |