

## DATA

The following data is from the articles published on arXiv in the Astrophysics domain. You will find 3 files: Data\_Train.csv, Data\_Test.csv and Submit.csv.

Data\_Train.csv and Data\_Test.csv have the following fields:

Title: The title of the article.

Date: The date the article was first posted to the arXiv.

In Data\_Train.csv the dates are all before May 2019, and in Data\_Test.csv the dates are all greater than or equal to May 1, 2019.

Abstract: The abstract of the article.

Subjects: The subject of the article, as described by the author. This field is only present in Data\_Train.csv.

## QUESTIONS

Q 1: Predict the number of articles in each subject for articles in Data\_Test.csv.

Predict the number of articles in each subject, in every month of Data\_Test.csv.

Q 2: Describe in 5 Bullet Points

- 1) Which algorithms did you try and why?
- 2) How did you measure the performance of your algorithms.
- 3) What else would you do if you had another day / week?
- 4) Your estimated 'score' on the test set.

You will be evaluated on:

- 1) Returning the assignment on time.
- 2) The performance of your model.
- 3) How well you describe your methods and future work.

Your returned assignment should include:

1. A working code for the problem above.
2. A filled submission file (Submit.csv). Replace the zeros in this file with your predicted values.

Do not include large files (e.g. the original data) with your submission.