

[Competitions](#)[Hackathons](#)[Data Scientists](#)[Discussions](#)[Jobs](#)[Learning](#)

# AI4D Swahili News Classification Challenge

## \$250 USD

Can you create a classification algorithm to identify Swahili news articles by category?

74 data scientists enrolled, 39 on the leaderboard

[Media](#) [Classification](#) [NLP](#)

East Africa

26 February 07:00—28 February 15:00

Ends in 20 hours

[Info](#)[Data](#)[Discussions](#)[Leaderboard](#)[Team](#)[Submissions](#)[Submit](#)

[train.csv](#) (49.9 MB)

[test.csv](#) (16.3 MB)

[sample\\_submission.csv](#) (111 KB)

StarterNotebook.ipynb (19.8 KB)

VariableDefinitions.csv (157 B)

The dataset describes 31,024 rows of news from different sources (most are from Tanzania). These news are in 6 different news categories from national news to entertainment news.

Your goal is to accurately classify each Swahili news content into six specified categories below:

- Kitaifa (National)
- Kimataifa (International)
- Uchumi (Business/Economy)
- Afya (Health)
- Michezo (Sports)
- Burudani (Entertainment)

### Variable definitions

- **id** - This is the id of particular news
- **content** - This is the content of particular news
- **category** - This is a category for particular news among five categories identified.

### The files for download

- **train.csv** is the dataset that you will use to train your model. This dataset includes 23,268 randomly selected news headlines.
- **test.csv** is the dataset to which you will apply your model to test how well it performs. Use your model and this dataset to predict in which of the five

categories the content of the particular news will be categorized. The test set contains 7,756 news headlines. This dataset includes the same fields as train.csv *except* for the last column. Note that the target is category.

- **sample\_submission.csv** is an example of what your submission file should look like.
- **StarterNotebook.ipynb** - this notebook will help you read in the data, build a

[Competitions](#)[Hackathons](#)[Data Scientists](#)[Discussions](#)[Jobs Board](#)[Host competition](#)[About Us](#)[Our Partners](#)[Contact Us](#)[Rules and Guidelines](#)[Terms of Use](#)[Privacy Policy](#)[FAQs](#)[LinkedIn](#)[Facebook](#)[Twitter](#)[Instagram](#)[Medium](#)[Youtube](#)[Github](#)

