Goal: Create a NLP model to fact-check articles on current events.

Packages used: Pandas, TF-IDF Vectorizer, XGBoost, Catboost

First approach: The data is defined into 6 classes: true, mostly-true, half-true, barely-true, false, not known. The data distribution is as follows.

* Half-True – 2
* False – 1
* Mostly-True – 3
* True – 5
* Barely-True – 0
* Not-Known - 4

A graph of blue bars

Description automatically generated

To Do:

SMOTE: Generate synthetic samples for the minority class in an imbalanced dataset.