The TextMine working group aims to bring together researchers on the broad topic of
broad theme of text mining. The TextMine challenge arises from a problem specific to text
(in this case, the detection of geographical named entities), the solution is provided in the form
in the form of a challenge, by making previously unpublished datasets available to the
scientific community. This is the context in which we are taking part in the 2024 edition: the problem is posed
by the Institut national de l'information géographique et forestière (IGN) and the Service hydro-
graphique et océanographique de la Marine (SHO).
and Oceanographic Service (Shom). The aim is to recognize geographical entities
in a corpus of nautical instructions. To meet this challenge, participants
participants have at their disposal a corpus made up of extracts from 15 volumes of nautical
nautical instructions annotated according to three distinct labels: name, for pure proper nouns ;
geogFeat, for common nouns that identify a geographical feature; and
geogName, for names associated with a geographical feature. The dataset is
66030 tokens and 18537 labels, and each token can have up to two labels.
The dataset is available on the Kaggle platform at this link 1. To evaluate the scores
scores, the model must predict the entities named in the unannotated test.csv file,
and participants are ranked according to the resulting micro f-measure.