

Sudan Crisis 2

July 13, 2025

Human Rights Monitoring and Protection of Population in Sudan Using Twitter Data

```
[1]: !java -version
```

```
openjdk version "11.0.27" 2025-04-15
OpenJDK Runtime Environment (build 11.0.27+6-post-Ubuntu-0ubuntu122.04)
OpenJDK 64-Bit Server VM (build 11.0.27+6-post-Ubuntu-0ubuntu122.04, mixed mode,
sharing)
```

```
[2]: %pip install kafka-python pandas geopandas matplotlib seaborn nltk spacy torch
↳ transformers pyspark findspark scikit-learn shapely geopy tensorflow==2.15.0
↳ tensorflow-hub==0.15.0 tensorflow-text==2.15.0 keras==2.15.0 numpy
```

```
Requirement already satisfied: kafka-python in
/home/hadoop/py311env/lib/python3.11/site-packages (2.2.10)
Requirement already satisfied: pandas in
/home/hadoop/py311env/lib/python3.11/site-packages (2.2.3)
Requirement already satisfied: geopandas in
/home/hadoop/py311env/lib/python3.11/site-packages (1.0.1)
Requirement already satisfied: matplotlib in
/home/hadoop/py311env/lib/python3.11/site-packages (3.10.3)
Requirement already satisfied: seaborn in
/home/hadoop/py311env/lib/python3.11/site-packages (0.13.2)
Requirement already satisfied: nltk in
/home/hadoop/py311env/lib/python3.11/site-packages (3.9.1)
Requirement already satisfied: spacy in
/home/hadoop/py311env/lib/python3.11/site-packages (3.8.7)
Requirement already satisfied: torch in
/home/hadoop/py311env/lib/python3.11/site-packages (2.7.0)
Requirement already satisfied: transformers in
/home/hadoop/py311env/lib/python3.11/site-packages (4.53.1)
Requirement already satisfied: pyspark in
/home/hadoop/py311env/lib/python3.11/site-packages (4.0.0)
Requirement already satisfied: findspark in
/home/hadoop/py311env/lib/python3.11/site-packages (2.0.1)
Requirement already satisfied: scikit-learn in
/home/hadoop/py311env/lib/python3.11/site-packages (1.6.1)
Requirement already satisfied: shapely in
/home/hadoop/py311env/lib/python3.11/site-packages (2.1.1)
```

Requirement already satisfied: geopy in
/home/hadoop/py311env/lib/python3.11/site-packages (2.4.1)

Requirement already satisfied: tensorflow==2.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (2.15.0)

Requirement already satisfied: tensorflow-hub==0.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (0.15.0)

Requirement already satisfied: tensorflow-text==2.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (2.15.0)

Requirement already satisfied: keras==2.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (2.15.0)

Requirement already satisfied: numpy in
/home/hadoop/py311env/lib/python3.11/site-packages (1.26.4)

Requirement already satisfied: absl-py>=1.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(2.3.0)

Requirement already satisfied: astunparse>=1.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(1.6.3)

Requirement already satisfied: flatbuffers>=23.5.26 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(25.2.10)

Requirement already satisfied: gast!=0.5.0,!0.5.1,!0.5.2,>=0.2.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(0.6.0)

Requirement already satisfied: google-pasta>=0.1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(0.2.0)

Requirement already satisfied: h5py>=2.9.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(3.13.0)

Requirement already satisfied: libclang>=13.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(18.1.1)

Requirement already satisfied: ml-dtypes~=0.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(0.2.0)

Requirement already satisfied: opt-einsum>=2.3.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(3.4.0)

Requirement already satisfied: packaging in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(25.0)

Requirement already satisfied:
protobuf!=4.21.0,!4.21.1,!4.21.2,!4.21.3,!4.21.4,!4.21.5,<5.0.0dev,>=3.20.3
in /home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(4.25.8)

Requirement already satisfied: setuptools in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)

(65.5.0)

Requirement already satisfied: six>=1.12.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(1.17.0)

Requirement already satisfied: termcolor>=1.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(3.1.0)

Requirement already satisfied: typing-extensions>=3.6.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(4.13.2)

Requirement already satisfied: wrapt<1.15,>=1.11.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(1.14.1)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(0.37.1)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(1.71.0)

Requirement already satisfied: tensorboard<2.16,>=2.15 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(2.15.2)

Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from tensorflow==2.15.0)
(2.15.0)

Requirement already satisfied: google-auth<3,>=1.6.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
tensorboard<2.16,>=2.15->tensorflow==2.15.0) (2.40.3)

Requirement already satisfied: google-auth-oauthlib<2,>=0.5 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
tensorboard<2.16,>=2.15->tensorflow==2.15.0) (1.2.2)

Requirement already satisfied: markdown>=2.6.8 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
tensorboard<2.16,>=2.15->tensorflow==2.15.0) (3.8)

Requirement already satisfied: requests<3,>=2.21.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
tensorboard<2.16,>=2.15->tensorflow==2.15.0) (2.32.3)

Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
tensorboard<2.16,>=2.15->tensorflow==2.15.0) (0.7.2)

Requirement already satisfied: werkzeug>=1.0.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
tensorboard<2.16,>=2.15->tensorflow==2.15.0) (3.1.3)

Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from google-
auth<3,>=1.6.3->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (5.5.2)

Requirement already satisfied: pyasn1-modules>=0.2.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from google-

auth<3,>=1.6.3->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (0.4.2)
 Requirement already satisfied: rsa<5,>=3.1.4 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from google-
 auth<3,>=1.6.3->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (4.9.1)
 Requirement already satisfied: requests-oauthlib>=0.7.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from google-auth-
 oauthlib<2,>=0.5->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (2.0.0)
 Requirement already satisfied: charset-normalizer<4,>=2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3,>=2.21.0->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (3.4.2)
 Requirement already satisfied: idna<4,>=2.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3,>=2.21.0->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (3.10)
 Requirement already satisfied: urllib3<3,>=1.21.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3,>=2.21.0->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (2.4.0)
 Requirement already satisfied: certifi>=2017.4.17 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 requests<3,>=2.21.0->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (2025.4.26)
 Requirement already satisfied: pyasn1>=0.1.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from rsa<5,>=3.1.4->google-
 auth<3,>=1.6.3->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (0.6.1)
 Requirement already satisfied: python-dateutil>=2.8.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from pandas) (2.9.0.post0)
 Requirement already satisfied: pytz>=2020.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from pandas) (2025.2)
 Requirement already satisfied: tzdata>=2022.7 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from pandas) (2025.2)
 Requirement already satisfied: pyogrio>=0.7.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from geopandas) (0.11.0)
 Requirement already satisfied: pyproj>=3.3.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from geopandas) (3.7.1)
 Requirement already satisfied: contourpy>=1.0.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (1.3.2)
 Requirement already satisfied: cycycler>=0.10 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (0.12.1)
 Requirement already satisfied: fonttools>=4.22.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (4.58.1)
 Requirement already satisfied: kiwisolver>=1.3.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (1.4.8)
 Requirement already satisfied: pillow>=8 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (11.2.1)
 Requirement already satisfied: pyparsing>=2.3.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from matplotlib) (3.2.3)
 Requirement already satisfied: click in
 /home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (8.2.1)
 Requirement already satisfied: joblib in
 /home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (1.5.1)

Requirement already satisfied: regex>=2021.8.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (2024.11.6)

Requirement already satisfied: tqdm in
/home/hadoop/py311env/lib/python3.11/site-packages (from nltk) (4.67.1)

Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.0.12)

Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (1.0.5)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (1.0.13)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.0.11)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.0.10)

Requirement already satisfied: thinc<8.4.0,>=8.3.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (8.3.4)

Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (1.1.3)

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.5.1)

Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.0.10)

Requirement already satisfied: weasel<0.5.0,>=0.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (0.4.1)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (0.16.0)

Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (2.11.5)

Requirement already satisfied: jinja2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.1.6)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from spacy) (3.5.0)

Requirement already satisfied: language-data>=1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
langcodes<4.0.0,>=3.2.0->spacy) (1.3.0)

Requirement already satisfied: annotated-types>=0.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.7.0)

Requirement already satisfied: pydantic-core==2.33.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (2.33.2)

Requirement already satisfied: typing-inspection>=0.4.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.4.1)

Requirement already satisfied: blis<1.3.0,>=1.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
thinc<8.4.0,>=8.3.4->spacy) (1.2.1)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in

```

/home/hadoop/py311env/lib/python3.11/site-packages (from
thinc<8.4.0,>=8.3.4->spacy) (0.1.5)
Requirement already satisfied: shellingham>=1.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (14.0.0)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (0.21.1)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (7.1.0)
Requirement already satisfied: filelock in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (3.18.0)
Requirement already satisfied: sympy>=1.13.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (1.14.0)
Requirement already satisfied: networkx in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (3.5)
Requirement already satisfied: fsspec in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (2025.3.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.80)
Requirement already satisfied: nvidia-cudnn-cu12==9.5.1.17 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (9.5.1.17)
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.4.1)
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (11.3.0.4)
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (10.3.7.77)
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (11.7.1.2)
Requirement already satisfied: nvidia-cusparselt-cu12==12.5.4.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.5.4.2)
Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.5.4.2)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (0.6.3)
Requirement already satisfied: nvidia-nccl-cu12==2.26.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (2.26.2)
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (12.6.85)

```

Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (1.11.1.6)

Requirement already satisfied: triton==3.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from torch) (3.3.0)

Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.32.3)

Requirement already satisfied: pyyaml>=5.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (6.0.2)

Requirement already satisfied: tokenizers<0.22,>=0.21 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.21.1)

Requirement already satisfied: safetensors>=0.4.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.5.3)

Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from huggingface-hub<1.0,>=0.30.0->transformers) (1.1.2)

Requirement already satisfied: py4j==0.10.9.9 in
/home/hadoop/py311env/lib/python3.11/site-packages (from pyspark) (0.10.9.9)

Requirement already satisfied: scipy>=1.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from scikit-learn) (1.15.3)

Requirement already satisfied: threadpoolctl>=3.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from scikit-learn) (3.6.0)

Requirement already satisfied: geographiclib<3,>=1.52 in
/home/hadoop/py311env/lib/python3.11/site-packages (from geopy) (2.0)

Requirement already satisfied: wheel<1.0,>=0.23.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from astunparse>=1.6.0->tensorflow==2.15.0) (0.45.1)

Requirement already satisfied: marisa-trie>=1.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from language-data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.1)

Requirement already satisfied: oauthlib>=3.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<2,>=0.5->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (3.3.1)

Requirement already satisfied: markdown-it-py>=2.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.19.1)

Requirement already satisfied: mdurl~=0.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)

Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from sympy>=1.13.3->torch) (1.3.0)

Requirement already satisfied: MarkupSafe>=2.1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from werkzeug>=1.0.1->tensorboard<2.16,>=2.15->tensorflow==2.15.0) (3.0.2)

Note: you may need to restart the kernel to use updated packages.

```
[3]: import tensorflow as tf
import tensorflow_hub as hub
import tensorflow_text as text

print("TensorFlow:", tf.__version__)
print("TensorFlow Hub:", hub.__version__)
print("TensorFlow Text:", text.__version__)
```

```
2025-07-07 17:33:53.784122: E
external/local_xla/xla/stream_executor/cuda/cuda_dnn.cc:9261] Unable to register
cuDNN factory: Attempting to register factory for plugin cuDNN when one has
already been registered
2025-07-07 17:33:53.784191: E
external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:607] Unable to register
cuFFT factory: Attempting to register factory for plugin cuFFT when one has
already been registered
2025-07-07 17:33:53.796881: E
external/local_xla/xla/stream_executor/cuda/cuda_blas.cc:1515] Unable to
register cuBLAS factory: Attempting to register factory for plugin cuBLAS when
one has already been registered
2025-07-07 17:33:53.849274: I tensorflow/core/platform/cpu_feature_guard.cc:182]
This TensorFlow binary is optimized to use available CPU instructions in
performance-critical operations.
To enable the following instructions: AVX2 FMA, in other operations, rebuild
TensorFlow with the appropriate compiler flags.
2025-07-07 17:33:54.818070: W
tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not
find TensorRT

TensorFlow: 2.15.0
TensorFlow Hub: 0.15.0
TensorFlow Text: 2.15.0
```

```
[4]: %pip install --upgrade jupyter ipywidgets
```

```
Requirement already satisfied: jupyter in
/home/hadoop/py311env/lib/python3.11/site-packages (1.1.1)
Requirement already satisfied: ipywidgets in
/home/hadoop/py311env/lib/python3.11/site-packages (8.1.7)
Requirement already satisfied: notebook in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (7.4.3)
Requirement already satisfied: jupyter-console in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (6.6.3)
Requirement already satisfied: nbconvert in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (7.16.6)
Requirement already satisfied: ipykernel in
```


/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (6.29.5)
 Requirement already satisfied: jupyterlab in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter) (4.4.3)
 Requirement already satisfied: comm>=0.1.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (0.2.2)
 Requirement already satisfied: ipython>=6.1.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (9.2.0)
 Requirement already satisfied: traitlets>=4.3.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (5.14.3)
 Requirement already satisfied: widgetsnbextension~=4.0.14 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (4.0.14)
 Requirement already satisfied: jupyterlab_widgets~=3.0.15 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipywidgets) (3.0.15)
 Requirement already satisfied: decorator in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (5.2.1)
 Requirement already satisfied: ipython-pygments-lexers in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (1.1.1)
 Requirement already satisfied: jedi>=0.16 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (0.19.2)
 Requirement already satisfied: matplotlib-inline in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (0.1.7)
 Requirement already satisfied: pexpect>4.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (4.9.0)
 Requirement already satisfied: prompt_toolkit<3.1.0,>=3.0.41 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (3.0.51)
 Requirement already satisfied: pygments>=2.4.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (2.19.1)
 Requirement already satisfied: stack_data in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (0.6.3)
 Requirement already satisfied: typing_extensions>=4.6 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 ipython>=6.1.0->ipywidgets) (4.13.2)
 Requirement already satisfied: wcwidth in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 prompt_toolkit<3.1.0,>=3.0.41->ipython>=6.1.0->ipywidgets) (0.2.13)
 Requirement already satisfied: parso<0.9.0,>=0.8.4 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from
 jedi>=0.16->ipython>=6.1.0->ipywidgets) (0.8.4)
 Requirement already satisfied: ptyprocess>=0.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from

pexpect>4.3->ipython>=6.1.0->ipywidgets) (0.7.0)
 Requirement already satisfied: debugpy>=1.6.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (1.8.14)
 Requirement already satisfied: jupyter-client>=6.1.12 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (8.6.3)
 Requirement already satisfied: jupyter-core!=5.0.*,>=4.12 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (5.8.1)
 Requirement already satisfied: nest-asyncio in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (1.6.0)
 Requirement already satisfied: packaging in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (25.0)
 Requirement already satisfied: psutil in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (7.0.0)
 Requirement already satisfied: pyzmq>=24 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (26.4.0)
 Requirement already satisfied: tornado>=6.1 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from ipykernel->jupyter)
 (6.5.1)
 Requirement already satisfied: python-dateutil>=2.8.2 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 client>=6.1.12->ipykernel->jupyter) (2.9.0.post0)
 Requirement already satisfied: platformdirs>=2.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
 core!=5.0.*,>=4.12->ipykernel->jupyter) (4.3.8)
 Requirement already satisfied: six>=1.5 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from python-
 dateutil>=2.8.2->jupyter-client>=6.1.12->ipykernel->jupyter) (1.17.0)
 Requirement already satisfied: async-lru>=1.0.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (2.0.5)
 Requirement already satisfied: httpx>=0.25.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (0.28.1)
 Requirement already satisfied: jinja2>=3.0.3 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (3.1.6)
 Requirement already satisfied: jupyter-lsp>=2.0.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
 (2.2.5)
 Requirement already satisfied: jupyter-server<3,>=2.4.0 in
 /home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)

(2.16.0)

Requirement already satisfied: jupyterlab-server<3,>=2.27.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
(2.27.3)

Requirement already satisfied: notebook-shim>=0.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
(0.2.4)

Requirement already satisfied: setuptools>=41.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab->jupyter)
(65.5.0)

Requirement already satisfied: anyio>=3.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (4.9.0)

Requirement already satisfied: argon2-cffi>=21.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (23.1.0)

Requirement already satisfied: jupyter-events>=0.11.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (0.12.0)

Requirement already satisfied: jupyter-server-terminals>=0.4.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (0.5.3)

Requirement already satisfied: nbformat>=5.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (5.10.4)

Requirement already satisfied: overrides>=5.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (7.7.0)

Requirement already satisfied: prometheus-client>=0.9 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (0.22.0)

Requirement already satisfied: send2trash>=1.8.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.8.3)

Requirement already satisfied: terminado>=0.8.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (0.18.1)

Requirement already satisfied: websocket-client>=1.7 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.8.0)

Requirement already satisfied: babel>=2.10 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (2.17.0)

Requirement already satisfied: json5>=0.9.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (0.12.0)

Requirement already satisfied: jsonschema>=4.18.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-

server<3,>=2.27.1->jupyterlab->jupyter) (4.24.0)
Requirement already satisfied: requests>=2.31 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyterlab-
server<3,>=2.27.1->jupyterlab->jupyter) (2.32.3)
Requirement already satisfied: idna>=2.8 in
/home/hadoop/py311env/lib/python3.11/site-packages (from anyio>=3.1.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (3.10)
Requirement already satisfied: sniffio>=1.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from anyio>=3.1.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.3.1)
Requirement already satisfied: argon2-cffi-bindings in
/home/hadoop/py311env/lib/python3.11/site-packages (from
argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (21.2.0)
Requirement already satisfied: certifi in
/home/hadoop/py311env/lib/python3.11/site-packages (from
httpx>=0.25.0->jupyterlab->jupyter) (2025.4.26)
Requirement already satisfied: httpcore==1.* in
/home/hadoop/py311env/lib/python3.11/site-packages (from
httpx>=0.25.0->jupyterlab->jupyter) (1.0.9)
Requirement already satisfied: h11>=0.16 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
httpcore==1.*->httpx>=0.25.0->jupyterlab->jupyter) (0.16.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jinja2>=3.0.3->jupyterlab->jupyter) (3.0.2)
Requirement already satisfied: attrs>=22.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter)
(2025.4.1)
Requirement already satisfied: referencing>=0.28.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jsonschema>=4.18.0->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (0.25.1)
Requirement already satisfied: python-json-logger>=2.0.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (3.3.0)
Requirement already satisfied: pyyaml>=5.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (6.0.2)
Requirement already satisfied: rfc3339-validator in
/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (0.1.4)
Requirement already satisfied: rfc3986-validator>=0.1.1 in

```

/home/hadoop/py311env/lib/python3.11/site-packages (from jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (0.1.1)
Requirement already satisfied: fqdn in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.5.1)
Requirement already satisfied: isoduration in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (20.11.0)
Requirement already satisfied: jsonpointer>1.13 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (3.0.0)
Requirement already satisfied: uri-template in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.3.0)
Requirement already satisfied: webcolors>=24.6.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from jsonschema[format-
nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (24.11.1)
Requirement already satisfied: beautifulsoup4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(4.13.4)
Requirement already satisfied: bleach!=5.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
bleach[css]!=5.0.0->nbconvert->jupyter) (6.2.0)
Requirement already satisfied: defusedxml in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(0.7.1)
Requirement already satisfied: jupyterlab-pygments in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(0.3.0)
Requirement already satisfied: mistune<4,>=2.0.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(3.1.3)
Requirement already satisfied: nbclient>=0.5.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(0.10.2)
Requirement already satisfied: pandocfilters>=1.4.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from nbconvert->jupyter)
(1.5.1)
Requirement already satisfied: webencodings in
/home/hadoop/py311env/lib/python3.11/site-packages (from
bleach!=5.0.0->bleach[css]!=5.0.0->nbconvert->jupyter) (0.5.1)
Requirement already satisfied: tinycss2<1.5,>=1.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from

```

```

bleach[css] != 5.0.0->nbconvert->jupyter) (1.4.0)
Requirement already satisfied: fastjsonschema>=2.15 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
nbformat>=5.3.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (2.21.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
requests>=2.31->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (3.4.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
requests>=2.31->jupyterlab-server<3,>=2.27.1->jupyterlab->jupyter) (2.4.0)
Requirement already satisfied: cffi>=1.0.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from argon2-cffi-
bindings->argon2-cffi>=21.1->jupyter-server<3,>=2.4.0->jupyterlab->jupyter)
(1.17.1)
Requirement already satisfied: pycparser in
/home/hadoop/py311env/lib/python3.11/site-packages (from
cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi>=21.1->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (2.22)
Requirement already satisfied: soupsieve>1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
beautifulsoup4->nbconvert->jupyter) (2.7)
Requirement already satisfied: arrow>=0.15.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
isoduration->jschema[format-nongpl]>=4.18.0->jupyter-events>=0.11.0->jupyter-
server<3,>=2.4.0->jupyterlab->jupyter) (1.3.0)
Requirement already satisfied: types-python-dateutil>=2.8.10 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
arrow>=0.15.0->isoduration->jschema[format-nongpl]>=4.18.0->jupyter-
events>=0.11.0->jupyter-server<3,>=2.4.0->jupyterlab->jupyter) (2.9.0.20250516)
Requirement already satisfied: executing>=1.2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
stack_data->ipython>=6.1.0->ipywidgets) (2.2.0)
Requirement already satisfied: asttokens>=2.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
stack_data->ipython>=6.1.0->ipywidgets) (3.0.0)
Requirement already satisfied: pure-eval in
/home/hadoop/py311env/lib/python3.11/site-packages (from
stack_data->ipython>=6.1.0->ipywidgets) (0.2.3)
Note: you may need to restart the kernel to use updated packages.

```

```
[5]: !jupyter nbextension enable --py widgetsnbextension
```

```

usage: jupyter [-h] [--version] [--config-dir] [--data-dir] [--runtime-dir]
               [--paths] [--json] [--debug]
               [subcommand]

```

Jupyter: Interactive Computing

positional arguments:

subcommand the subcommand to launch

options:

-h, --help show this help message and exit
--version show the versions of core jupyter packages and exit
--config-dir show Jupyter config dir
--data-dir show Jupyter data dir
--runtime-dir show Jupyter runtime dir
--paths show all Jupyter paths. Add --json for machine-readable format.
--json output paths as machine-readable json
--debug output debug information about paths

Available subcommands: console dejavu events execute kernel kernelspec lab labextension labhub migrate nbconvert notebook run server troubleshoot trust

Jupyter command `jupyter-nbextension` not found.

```
[6]: import sys
      print(sys.executable)
```

/home/hadoop/py311env/bin/python

```
[7]: import os
      os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-11-openjdk-amd64"
      os.environ["SPARK_HOME"] = "/opt/spark/spark-3.5.6-bin-hadoop3"
      os.environ["HADOOP_HOME"] = "/usr/local/hadoop"
      os.environ["TF_USE_LEGACY_KERAS"] = "1"
```

```
[8]: import findspark
      findspark.init()
      import pyspark
```

```
[9]: #Import necessary libraries
      import pandas as pd # Pandas is an open source, BSD-licensed library
      ↪providing high-performance, easy-to-use data structures and data analysis
      ↪tools for the Python programming language. Pandas Series is a
      ↪one-dimensional labeled array capable of holding any data type with axis
      ↪labels or index
      import geopandas as gpd
      import time
      import matplotlib.pyplot as plt
      import numpy as np
      import logging
      import hashlib
      from geopy.geocoders import Nominatim
```

```

from geopy.extra.rate_limiter import RateLimiter
from pyspark.sql import functions as func
from pyspark.sql.types import IntegerType, StringType, FloatType,
    ↳TimestampType, BinaryType
from pyspark.sql.functions import col, count, when, mean, stddev, abs, sum,
    ↳row_number
from pyspark.ml.feature import VectorAssembler, StandardScaler, StringIndexer,
    ↳OneHotEncoder
from pyspark.ml import Pipeline
from pyspark.ml.classification import RandomForestClassifier
from pyspark.ml.tuning import CrossValidator, ParamGridBuilder
from pyspark.ml.evaluation import MulticlassClassificationEvaluator,
    ↳BinaryClassificationEvaluator
from sklearn.metrics import precision_recall_curve
from pyspark.sql import Window
from transformers import pipeline, AutoTokenizer,
    ↳AutoModelForSequenceClassification
from shapely.geometry import Point
import torch

```

```

[10]: import nltk
nltk.download('punkt')    # Download NLTK essential resources
!python -m spacy download en_core_web_sm    # Download spaCy Language Model

```

[nltk_data] Downloading package punkt to /home/hadoop/nltk_data...

[nltk_data] Package punkt is already up-to-date!

Collecting en-core-web-sm==3.8.0

Downloading https://github.com/explosion/spacy-
models/releases/download/en_core_web_sm-3.8.0/en_core_web_sm-3.8.0-py3-none-
any.whl (12.8 MB)

12.8/12.8 MB

36.1 MB/s eta 0:00:00 0:00:01

Download and installation successful

You can now load the package via spacy.load('en_core_web_sm')

```

[11]: import nltk
nltk.download('punkt_tab')

```

[nltk_data] Downloading package punkt_tab to /home/hadoop/nltk_data...

[nltk_data] Package punkt_tab is already up-to-date!

[11]: True

```

[12]: # Test NLTK installation
from nltk.tokenize import word_tokenize
text = "Hello, world!"

```



```
tokens = word_tokenize(text)
print(tokens)
```

```
['Hello', ',', 'world', '!']
```

```
[13]: import spacy
      try:
          nlp = spacy.load("en_core_web_sm")
      except OSError:
          from spacy.cli import download
          download("en_core_web_sm")
          nlp = spacy.load("en_core_web_sm")

      # Test spaCy installation
      doc = nlp("Hello, world!")
      print([token.text for token in doc])
```

```
['Hello', ',', 'world', '!']
```

```
[14]: # Start SparkSession
      from pyspark.sql import SparkSession
      spark=SparkSession.builder.appName("Sudan Crisis Analysis").config("spark.
          ↪driver.memory", "4g").getOrCreate()
      spark
```

```
25/07/07 17:34:21 WARN Utils: Your hostname, julius-virtual-machine resolves to
a loopback address: 127.0.1.1; using 192.168.145.130 instead (on interface
ens33)
```

```
25/07/07 17:34:21 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
```

```
Setting default log level to "WARN".
```

```
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use
setLogLevel(newLevel).
```

```
25/07/07 17:34:22 WARN NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
```

```
[14]: <pyspark.sql.session.SparkSession at 0x710b0870abd0>
```

```
[15]: #Load the data and handle missing values
      df_pyspark = spark.read.csv("Sudan_Twitter_Data_June_2025latlong.csv",
          ↪header=True, inferSchema=True).cache()
      df_pyspark.show(5)
```

```
25/07/07 17:34:30 WARN SparkStringUtils: Truncated the string representation of
a plan since it was too large. This behavior can be adjusted by setting
'spark.sql.debug.maxToStringFields'.
```

```
[Stage 2:=====>
```

```
(1 + 2) / 3]
```



```

02A4D3|https://pbs.twimg...|2019-12-22T17:21:...|          NULL|          14|
204|    0|          2549|    FALSE|          NULL|  en|retweet|RT @BSonblast:
Pl...|2025-06-17T03:29:...|  "["@bsonblast"]"|          0|          0|          0|
[]|    NULL|    []|          0|    FALSE|    NULL|Abyei
PCA|https://x.com/use...|          NULL|    FALSE|    NULL|          NULL|
NULL|9.594021206|28.43772416|
+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+
only showing top 5 rows

```

```

[16]: # Convert PySpark DataFrame to Pandas
filtered_df_pandas = df_pyspark.toPandas()

```

```

[17]: # Configure logging
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s -_
↳%(message)s')

BOT_THRESHOLD = 0.5
MEDIA_THRESHOLD = 0.7

def process_tweets(scraped_tweets):
    filtered_tweets = []
    total = 0
    skipped_bot = 0
    skipped_media = 0

    for tweet in scraped_tweets:
        total += 1
        try:
            # Bot detection step
            bot_score = bot_detection_model(tweet.user_metadata, tweet.text)
            if bot_score > BOT_THRESHOLD:
                skipped_bot += 1
                logging.info(f"Skipped tweet {tweet.id} due to high bot score:_
↳{bot_score}")
                continue

            # Media verification step (if media present)
            if tweet.has_media:
                media_credibility = verify_media(tweet.media_url)
                if media_credibility < MEDIA_THRESHOLD:

```

```

        skipped_media += 1
        logging.info(f"Skipped tweet {tweet.id} due to low media_
↳credibility: {media_credibility}")
        continue

    # Passed both checks
    filtered_tweets.append(tweet)
    logging.info(f"Accepted tweet {tweet.id} for further processing")

except Exception as e:
    logging.error(f"Error processing tweet {tweet.id}: {e}")
    continue

logging.info(f"Processed {total} tweets: {len(filtered_tweets)} accepted,
↳{skipped_bot} skipped (bot), {skipped_media} skipped (media)")
return filtered_tweets

```

```
[18]: print(filtered_df_pandas[['Latitude', 'Longitude']].head(100))
```

	Latitude	Longitude
0	9.594021206	28.43772416
1	9.594021206	28.43772416
2	9.594021206	28.43772416
3	9.594021206	28.43772416
4	9.594021206	28.43772416
..
95	9.594021206	28.43772416
96	9.594021206	28.43772416
97	9.594021206	28.43772416
98	9.594021206	28.43772416
99	9.594021206	28.43772416

[100 rows x 2 columns]

```
[19]: import pandas as pd

filtered_df_pandas['Longitude'] = pd.
↳to_numeric(filtered_df_pandas['Longitude'], errors='coerce')
filtered_df_pandas['Latitude'] = pd.to_numeric(filtered_df_pandas['Latitude'],
↳errors='coerce')
```

```
[20]: # Load Sudan states boundaries to create geolocation filtering pipeline
states_gdf = gpd.read_file('sudan_states.geojson')
```

```
[21]: # Create GeoDataFrame from filtered tweets
from shapely.geometry import Point
```

```

geometry = [Point(xy) for xy in zip(filtered_df_pandas['Longitude'],
↳filtered_df_pandas['Latitude'])]
tweets_gdf = gpd.GeoDataFrame(filtered_df_pandas, geometry=geometry, crs='EPSG:
↳4326')

```

```
[22]: print(states_gdf.columns)
```

```

Index(['Shape_Leng', 'Shape_Area', 'ADM1_EN', 'ADM1_AR', 'ADM1_PCODE',
      'ADM1_REF', 'ADM1ALT1EN', 'ADM1ALT2EN', 'ADM1ALT1AR', 'ADM1ALT2AR',
      'ADMO_EN', 'ADMO_AR', 'ADMO_PCODE', 'date', 'validOn', 'Incidents',
      'geometry'],
      dtype='object')

```

```
[23]: print('index_right' in tweets_gdf.columns)
print('index_right' in states_gdf.columns)
```

```

False
False

```

```
[24]: if 'index_right' in tweets_gdf.columns:
      tweets_gdf = tweets_gdf.drop(columns=['index_right'])
```

```
[25]: # Spatial join to assign state to each tweet
tweets_gdf = gpd.sjoin(
    tweets_gdf,
    states_gdf[['ADM1_EN', 'geometry']],
    how='left',
    predicate='within'
)

# Rename the column for clarity
tweets_gdf.rename(columns={'ADM1_EN': 'state'}, inplace=True)

# Check results
print(tweets_gdf[['state', 'geometry']].head())

```

```

      state      geometry
0  Abyei PCA  POINT (28.43772 9.59402)
1  Abyei PCA  POINT (28.43772 9.59402)
2  Abyei PCA  POINT (28.43772 9.59402)
3  Abyei PCA  POINT (28.43772 9.59402)
4  Abyei PCA  POINT (28.43772 9.59402)

```

```
[26]: from collections import Counter

col_counts = Counter(tweets_gdf.columns)
duplicates = [col for col, count in col_counts.items() if count > 1]
```

```
print("Duplicate columns:", duplicates)
```

Duplicate columns: []

```
[27]: tweets_gdf = tweets_gdf.loc[:, ~tweets_gdf.columns.duplicated()]
```

```
[28]: print(tweets_gdf.index.is_unique)
```

True

```
[29]: tweets_gdf.index = pd.RangeIndex(start=0, stop=len(tweets_gdf), step=1)
```

```
[30]: # Remove points with empty or invalid geometries and missing states
tweets_gdf_filtered = tweets_gdf[
    (~tweets_gdf.geometry.is_empty) &
    (tweets_gdf.geometry.is_valid) &
    (tweets_gdf['state'].notna())
].copy()

print(tweets_gdf_filtered[['state', 'geometry']].head())
```

	state	geometry
0	Abyei PCA	POINT (28.43772 9.59402)
1	Abyei PCA	POINT (28.43772 9.59402)
2	Abyei PCA	POINT (28.43772 9.59402)
3	Abyei PCA	POINT (28.43772 9.59402)
4	Abyei PCA	POINT (28.43772 9.59402)

```
[31]: print(tweets_gdf_filtered.head(10))
```

	_id	userId	userName	alias \
0	1.93E+18	2.437521e+08	Damascene	nanssolo
1	1.93E+18	1.450000e+18	nana nano	hydrowanter
2	1.93E+18	1.060000e+18	bean	Charlixxx_
3	1.93E+18	7.160000e+17	Pastel	soIguessIexist
4	1.93E+18	1.210000e+18	u i	02A4D3
5	1.93E+18	5.523299e+08	TGD	GrandFilmMaster
6	1.93E+18	7.060000e+17	WIP I Guess	ctdjgurl
7	1.93E+18	1.930000e+18	SGUY2004	SGUY20044
8	1.93E+18	1.540000e+18		us3rsnam3
9	1.93E+18	1.616353e+09	Superstar Christie In A Box	butt_um_liam

	profilePicture \
0	https://pbs.twimg.com/profile_images/173403408...
1	https://pbs.twimg.com/profile_images/189138425...
2	https://pbs.twimg.com/profile_images/145558722...
3	https://pbs.twimg.com/profile_images/153254988...
4	https://pbs.twimg.com/profile_images/190776009...

5 https://pbs.twimg.com/profile_images/316694971...
6 https://pbs.twimg.com/profile_images/185726903...
7 https://pbs.twimg.com/profile_images/192635389...
8 https://pbs.twimg.com/profile_images/181146951...
9 https://pbs.twimg.com/profile_images/175423406...

	userCreatedAt \
0	2011-01-27T18:51:26.000Z
1	2021-10-05T06:07:55.000Z
2	2018-11-13T12:11:55.000Z
3	2016-04-01T01:12:51.000Z
4	2019-12-22T17:21:09.000Z
5	2012-04-13T01:40:29.000Z
6	2016-03-04T15:19:36.000Z
7	2025-05-24T17:45:01.000Z
8	2022-06-16T15:28:37.000Z
9	2013-07-23T23:38:46.000Z

	bio	followers	following \
0	Syrian* blunt and very loud*Regrettably an att...	941	754
1	she/her Indonesia 24 Aroace Hoyo ga...	156	46
2	pan pirate go brr	152	327
3	24 black nonbinary lesbian she/they Ar...	488	554
4	None	14	204
5	28BlackHe/HimI do whatever	219	2277
6	I love my Black People. Not your mule or negro...	397	584
7	Fourth time making a Twitter account GOD PLEAS...	216	954
8	old username: bratzondablock	1290	1076
9	(not so) Clever musings.she/her	271	347

	lists	...	noteTweetText	inReplyTo	inReplyToId	inReplyToUserId	reportName \
0	2	...	None	FALSE	None	None	None
1	1	...	None	FALSE	None	None	None
2	0	...	None	FALSE	None	None	None
3	2	...	None	FALSE	None	None	None
4	0	...	None	FALSE	None	None	None
5	6	...	None	FALSE	None	None	None
6	0	...	None	FALSE	None	None	None
7	0	...	None	FALSE	None	None	None
8	1	...	None	FALSE	None	None	None
9	4	...	None	FALSE	None	None	None

	Latitude	Longitude	geometry	index_right	state
0	9.594021	28.437724	POINT (28.43772 9.59402)	0.0	Abyei PCA
1	9.594021	28.437724	POINT (28.43772 9.59402)	0.0	Abyei PCA
2	9.594021	28.437724	POINT (28.43772 9.59402)	0.0	Abyei PCA
3	9.594021	28.437724	POINT (28.43772 9.59402)	0.0	Abyei PCA
4	9.594021	28.437724	POINT (28.43772 9.59402)	0.0	Abyei PCA

```

5  9.594021  28.437724  POINT (28.43772 9.59402)      0.0  Abyei PCA
6  9.594021  28.437724  POINT (28.43772 9.59402)      0.0  Abyei PCA
7  9.594021  28.437724  POINT (28.43772 9.59402)      0.0  Abyei PCA
8  9.594021  28.437724  POINT (28.43772 9.59402)      0.0  Abyei PCA
9  9.594021  28.437724  POINT (28.43772 9.59402)      0.0  Abyei PCA

```

[10 rows x 39 columns]

```

[32]: tweets_gdf['content_type'] = (
        (tweets_gdf['isImage'].astype(str).str.lower() == 'true') &
        tweets_gdf['firstImage'].notnull()
    ).map({True: 'image', False: 'text'})

```

```
[ ]:
```

```
[ ]:
```

```

[33]: # Use transformers model that supports Arabic sentiment analysis
sentiment_pipeline = pipeline(
    "sentiment-analysis",
    model="CAMEL-Lab/bert-base-arabic-camelbert-da-sentiment"
)

def arabic_sentiment_analysis(text):
    if not isinstance(text, str) or not text.strip():
        return "neutral" # or any default label you prefer
    result = sentiment_pipeline(text)
    return result[0]['label']

```

Device set to use cpu

```

[34]: print(sentiment_pipeline("      !"))

```

```
[{'label': 'positive', 'score': 0.9809406399726868}]
```

```

[35]: # Sentiment analysis (pseudo-code)
tweets_gdf['sentiment'] = tweets_gdf['text'].apply(arabic_sentiment_analysis)

```

```

[36]: print(tweets_gdf.columns.tolist())

```

```

['_id', 'userId', 'userName', 'alias', 'profilePicture', 'userCreatedAt', 'bio',
'followers', 'following', 'lists', 'statusesCount', 'verified', 'userLocation',
'lang', 'type', 'text', 'createdAt', 'mentions', 'totalRetweets', 'favorites',
'replies', 'links', 'firstLink', 'images', 'impressions', 'isImage',
'firstImage', 'location', 'url', 'noteTweetText', 'inReplyTo', 'inReplyToId',
'inReplyToUserId', 'reportName', 'Latitude', 'Longitude', 'geometry',
'index_right', 'state', 'content_type', 'sentiment']

```


[]:

[37]: %**pip** install --upgrade transformers "accelerate>=0.26.0"

huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid deadlocks...

To disable this warning, you can either:

- Avoid using `tokenizers` before the fork if possible
- Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)

IOStream.flush timed out

Requirement already satisfied: transformers in

/home/hadoop/py311env/lib/python3.11/site-packages (4.53.1)

Requirement already satisfied: accelerate>=0.26.0 in

/home/hadoop/py311env/lib/python3.11/site-packages (1.8.1)

Requirement already satisfied: filelock in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (3.18.0)

Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.32.3)

Requirement already satisfied: numpy>=1.17 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (1.26.4)

Requirement already satisfied: packaging>=20.0 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (25.0)

Requirement already satisfied: pyyaml>=5.1 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (6.0.2)

Requirement already satisfied: regex!=2019.12.17 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers)

(2024.11.6)

Requirement already satisfied: requests in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (2.32.3)

Requirement already satisfied: tokenizers<0.22,>=0.21 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.21.1)

Requirement already satisfied: safetensors>=0.4.3 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (0.5.3)

Requirement already satisfied: tqdm>=4.27 in

/home/hadoop/py311env/lib/python3.11/site-packages (from transformers) (4.67.1)

Requirement already satisfied: fsspec>=2023.5.0 in

/home/hadoop/py311env/lib/python3.11/site-packages (from huggingface-hub<1.0,>=0.30.0->transformers) (2025.3.0)

Requirement already satisfied: typing-extensions>=3.7.4.3 in

/home/hadoop/py311env/lib/python3.11/site-packages (from huggingface-hub<1.0,>=0.30.0->transformers) (4.13.2)

Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in

/home/hadoop/py311env/lib/python3.11/site-packages (from huggingface-hub<1.0,>=0.30.0->transformers) (1.1.2)

Requirement already satisfied: psutil in

/home/hadoop/py311env/lib/python3.11/site-packages (from accelerate>=0.26.0)

(7.0.0)
Requirement already satisfied: torch>=2.0.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from accelerate>=0.26.0)
(2.7.0)
Requirement already satisfied: sympy>=1.13.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (1.14.0)
Requirement already satisfied: networkx in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (3.5)
Requirement already satisfied: jinja2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (3.1.6)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (12.6.77)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (12.6.77)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (12.6.80)
Requirement already satisfied: nvidia-cudnn-cu12==9.5.1.17 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (9.5.1.17)
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (12.6.4.1)
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (11.3.0.4)
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (10.3.7.77)
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (11.7.1.2)
Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (12.5.4.2)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.3 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (0.6.3)
Requirement already satisfied: nvidia-nccl-cu12==2.26.2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (2.26.2)
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in
/home/hadoop/py311env/lib/python3.11/site-packages (from

```

torch>=2.0.0->accelerate>=0.26.0) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (12.6.85)
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (1.11.1.6)
Requirement already satisfied: triton==3.3.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
torch>=2.0.0->accelerate>=0.26.0) (3.3.0)
Requirement already satisfied: setuptools>=40.8.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
triton==3.3.0->torch>=2.0.0->accelerate>=0.26.0) (65.5.0)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
sympy>=1.13.3->torch>=2.0.0->accelerate>=0.26.0) (1.3.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/hadoop/py311env/lib/python3.11/site-packages (from
jinja2->torch>=2.0.0->accelerate>=0.26.0) (3.0.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/hadoop/py311env/lib/python3.11/site-packages (from requests->transformers)
(3.4.2)
Requirement already satisfied: idna<4,>=2.5 in
/home/hadoop/py311env/lib/python3.11/site-packages (from requests->transformers)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/hadoop/py311env/lib/python3.11/site-packages (from requests->transformers)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/home/hadoop/py311env/lib/python3.11/site-packages (from requests->transformers)
(2025.4.26)
Note: you may need to restart the kernel to use updated packages.

```

```
[38]: import transformers
      print(transformers.__version__)
```

4.53.1

```
[39]: import accelerate
      print(accelerate.__version__)
```

1.8.1

```
[40]: # Sentiment Analysis with RoBERTa (Hugging Face Transformers)

import torch
from transformers import RobertaTokenizer, RobertaForSequenceClassification, \
    Trainer, TrainingArguments
```

```

import evaluate
from sklearn.model_selection import train_test_split
import pandas as pd
import numpy as np

# --- 1. Data Preparation and Cleaning ---
# Remove rows with missing text or sentiment
tweets_gdf = tweets_gdf.dropna(subset=['text', 'sentiment'])

# Map sentiment labels to integers
label2id = {'negative': 0, 'neutral': 1, 'positive': 2}
id2label = {v: k for k, v in label2id.items()}
tweets_gdf['label'] = tweets_gdf['sentiment'].map(label2id)

# Remove rows where mapping failed (unmapped sentiment)
tweets_gdf = tweets_gdf.dropna(subset=['label'])

# Ensure text is string type
tweets_gdf['text'] = tweets_gdf['text'].astype(str)

# --- 2. Train-validation Split ---
train_texts, val_texts, train_labels, val_labels = train_test_split(
    tweets_gdf['text'].tolist(),
    tweets_gdf['label'].astype(int).tolist(),
    test_size=0.2,
    random_state=42
)

# --- 3. Load Tokenizer and Model ---
model_name = "syedkhalid076/Roberta-Sentimental-Analysis-v2"
tokenizer = RobertaTokenizer.from_pretrained(model_name)
model = RobertaForSequenceClassification.from_pretrained(
    model_name,
    num_labels=3,
    id2label=id2label,
    label2id=label2id
)

# --- 4. Tokenize Data ---
train_encodings = tokenizer(train_texts, truncation=True, padding=True,
    ↪max_length=128)
val_encodings = tokenizer(val_texts, truncation=True, padding=True,
    ↪max_length=128)

# --- 5. Prepare Dataset Objects ---
class TweetDataset(torch.utils.data.Dataset):
    def __init__(self, encodings, labels):

```

```

        self.encodings = encodings
        self.labels = labels
    def __getitem__(self, idx):
        item = {k: torch.tensor(v[idx]) for k, v in self.encodings.items()}
        item['labels'] = torch.tensor(self.labels[idx])
        return item
    def __len__(self):
        return len(self.labels)

train_dataset = TweetDataset(train_encodings, train_labels)
val_dataset = TweetDataset(val_encodings, val_labels)

# --- 6. Define Trainer and Training Arguments ---
training_args = TrainingArguments(
    output_dir='./results',
    num_train_epochs=2,
    per_device_train_batch_size=8,
    per_device_eval_batch_size=8,
    eval_strategy="epoch",
    save_strategy="epoch",
    logging_dir='./logs',
    logging_steps=50,
    load_best_model_at_end=True,
    metric_for_best_model="accuracy"
)

metric = evaluate.load("accuracy")

def compute_metrics(eval_pred):
    logits, labels = eval_pred
    predictions = np.argmax(logits, axis=-1)
    return metric.compute(predictions=predictions, references=labels)

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=val_dataset,
    compute_metrics=compute_metrics,
    tokenizer=tokenizer
)

# --- 7. Train the Model ---
trainer.train()

# --- 8. Evaluate the Model ---
eval_results = trainer.evaluate()

```

```
print("Evaluation results:", eval_results)
```

```
2025-07-07 18:17:25,178 - INFO - PyTorch version 2.7.0 available.
2025-07-07 18:17:25,180 - INFO - TensorFlow version 2.15.0 available.
/tmp/ipykernel_6568/562764098.py:83: FutureWarning: `tokenizer` is deprecated
and will be removed in version 5.0.0 for `Trainer.__init__`. Use
`processing_class` instead.
```

```
    trainer = Trainer(
2025-07-07 18:17:33,794 - INFO - Error while sending or receiving.
Traceback (most recent call last):
  File "/opt/spark/spark-3.5.6-bin-
hadoop3/python/lib/py4j-0.10.9.7-src.zip/py4j/clientserver.py", line 503, in
send_command
```

```
    self.socket.sendall(command.encode("utf-8"))
ConnectionResetError: [Errno 104] Connection reset by peer
2025-07-07 18:17:33,797 - INFO - Closing down clientserver connection
2025-07-07 18:17:33,800 - INFO - Exception while sending command.
```

```
Traceback (most recent call last):
  File "/opt/spark/spark-3.5.6-bin-
hadoop3/python/lib/py4j-0.10.9.7-src.zip/py4j/clientserver.py", line 503, in
send_command
```

```
    self.socket.sendall(command.encode("utf-8"))
ConnectionResetError: [Errno 104] Connection reset by peer
```

During handling of the above exception, another exception occurred:

```
Traceback (most recent call last):
  File "/opt/spark/spark-3.5.6-bin-
hadoop3/python/lib/py4j-0.10.9.7-src.zip/py4j/java_gateway.py", line 1038, in
send_command
```

```
    response = connection.send_command(command)
    ~~~~~
  File "/opt/spark/spark-3.5.6-bin-
hadoop3/python/lib/py4j-0.10.9.7-src.zip/py4j/clientserver.py", line 506, in
send_command
    raise Py4JNetworkError(
py4j.protocol.Py4JNetworkError: Error while sending
/home/hadoop/py311env/lib/python3.11/site-
packages/torch/utils/data/dataloader.py:665: UserWarning: 'pin_memory' argument
is set as true but no accelerator is found, then device pinned memory won't be
used.
```

```
    warnings.warn(warn_msg)
2025-07-07 18:17:33,806 - INFO - Closing down clientserver connection
```

<IPython.core.display.HTML object>

```
/home/hadoop/py311env/lib/python3.11/site-
packages/torch/utils/data/dataloader.py:665: UserWarning: 'pin_memory' argument
```

is set as true but no accelerator is found, then device pinned memory won't be used.

```
warnings.warn(warn_msg)
/home/hadoop/py311env/lib/python3.11/site-
packages/torch/utils/data/dataloader.py:665: UserWarning: 'pin_memory' argument
is set as true but no accelerator is found, then device pinned memory won't be
used.
```

```
warnings.warn(warn_msg)
```

<IPython.core.display.HTML object>

```
Evaluation results: {'eval_loss': 2.1457651655509835e-06, 'eval_accuracy': 1.0,
'eval_runtime': 374.5909, 'eval_samples_per_second': 8.348,
'eval_steps_per_second': 1.044, 'epoch': 2.0}
```

```
[41]: # Predict on the validation set
predictions_output = trainer.predict(val_dataset)
predicted_labels = np.argmax(predictions_output.predictions, axis=1)
predicted_sentiments = [id2label[label] for label in predicted_labels]

# Create DataFrame for export
val_df = pd.DataFrame({
    'text': val_texts,
    'true_label': val_labels,
    'predicted_label': predicted_labels,
    'predicted_sentiment': predicted_sentiments
})

# Export to CSV
val_df.to_csv("sentiment_analysis_predictions.csv", index=False)
print("Exported predictions to sentiment_analysis_predictions.csv")
```

```
/home/hadoop/py311env/lib/python3.11/site-
packages/torch/utils/data/dataloader.py:665: UserWarning: 'pin_memory' argument
is set as true but no accelerator is found, then device pinned memory won't be
used.
```

```
warnings.warn(warn_msg)
```

Exported predictions to sentiment_analysis_predictions.csv

```
[ ]:
```

```
[42]: # Load population and conflict severity data
pop_df = pd.read_csv('sudan_population_by_state.csv') # columns: state,␣
    ↳ population source: UNFPA 2024 COD-PS Update
conflict_df = pd.read_csv('sudan_conflict_severity.csv') # columns: state,␣
    ↳ severity_score source: ACLED

# Merge auxiliary data
```

```
weights_df = pop_df.merge(conflict_df, on='state')
```

```
[43]: # Remove commas robustly, even if some values are already numeric
weights_df['population'] = pd.to_numeric(weights_df['population'],
    ↪errors='coerce')
weights_df['severity_score'] = pd.to_numeric(weights_df['severity_score'],
    ↪errors='coerce')
```

```
[44]: print(weights_df[['state', 'population', 'severity_score']])
```

	state	population	severity_score
0	Aj Jazirah	5124749	4.03
1	Blue Nile	813930	0.29
2	Central Darfur	943721	0.50
3	East Darfur	2390080	0.16
4	Gedaref	3091393	0.06
5	Kassala	2718540	0.00
6	Khartoum	5941286	10.00
7	North Darfur	3461818	9.65
8	North Kordofan	3444769	2.16
9	Northern	1446861	0.07
10	Red Sea	2035582	0.00
11	River Nile	1862303	0.12
12	Sennar	2532326	0.81
13	South Darfur	3895007	2.15
14	South Kordofan	1156469	1.58
15	West Darfur	618178	4.80
16	West Kordofan	2612654	1.52
17	White Nile	3392274	0.61

```
[45]: # Calculate combined weights
alpha = 0.5
weights_df['weight'] = (alpha * weights_df['population'] /
    ↪weights_df['population'].sum() +
                        (1 - alpha) * weights_df['severity_score'] /
    ↪weights_df['severity_score'].sum())

# Normalize weights
weights_df['weight'] /= weights_df['weight'].sum()

# Determine sample size per state
total_sample_size = 100
weights_df['sample_quota'] = (weights_df['weight'] * total_sample_size).round().
    ↪astype(int)
```

```
[46]: # For each state, allocate quotas by content_type and sentiment proportionally
final_samples = []
```



```

for state, quota in zip(weights_df['state'], weights_df['sample_quota']):
    state_tweets = tweets_gdf[tweets_gdf['state'] == state]
    # Calculate proportions of content_type and sentiment
    proportions = (state_tweets.groupby(['content_type', 'sentiment']).size() /
    ↪len(state_tweets)).reset_index(name='prop')
    for _, row in proportions.iterrows():
        n_samples = int(row['prop'] * quota)
        subset = state_tweets[(state_tweets['content_type'] ==
    ↪row['content_type']) &
                                (state_tweets['sentiment'] == row['sentiment'])]
        sampled_subset = subset.sample(n=min(n_samples, len(subset)),
    ↪random_state=42)
        final_samples.append(sampled_subset)

# Concatenate final stratified sample
final_sample_df = pd.concat(final_samples)

```

[47]: *#Pseudonymization: Replace Usernames with SHA-256 Hashes*
Secret pepper value (store securely)
 PEPPER = b"your-secret-pepper-value"

```

def pseudonymize_username(username):
    if not username:
        return None
    # Encode username + pepper as bytes
    data = f"{username}".encode() + PEPPER
    # Generate SHA-256 hash
    return hashlib.sha256(data).hexdigest()

# Apply to DataFrame
final_sample_df['userName'] = final_sample_df['userName'].
    ↪apply(pseudonymize_username)

```

[48]: *#Redaction: Strip GPS Coordinates to City-Level Precision*

```

# Configure geocoder with user agent
geolocator = Nominatim(user_agent="sudan_crisis_monitor_v1")

# Add rate limiting (1 request/sec)
reverse = RateLimiter(geolocator.reverse, min_delay_seconds=1)

def get_city_name(lat, lon):
    if pd.isna(lat) or pd.isna(lon):
        return None
    try:
        location = reverse((lat, lon), exactly_one=True, language="en")
        if location:

```

```

        address = location.raw.get('address', {})
        return address.get('city') or address.get('town') or address.
↳get('village')
    except Exception as e:
        logging.error(f"Geocoding failed for ({lat}, {lon}): {e}")
    return None

# Apply to DataFrame
final_sample_df['Location'] = final_sample_df.apply(lambda row:↳
↳get_city_name(row['Latitude'], row['Longitude']), axis=1)

# Remove precise coordinates
final_sample_df = final_sample_df.drop(columns=['Longitude', 'Latitude'])

```

```

2025-07-07 22:02:59,407 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:01,061 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:03,109 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:05,068 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:06,263 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:07,493 - WARNING - RateLimiter caught an error, retrying (0/2
tries). Called with ((*((14.39582814, 33.53808821),), **{'exactly_one': True,
'language': 'en'})).

```

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 534, in _make_request
    response = conn.getresponse()
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connection.py", line 516, in getresponse
    httplib_response = super().getresponse()
    ~~~~~
File "/usr/lib/python3.11/http/client.py", line 1395, in getresponse
    response.begin()
File "/usr/lib/python3.11/http/client.py", line 325, in begin
    version, status, reason = self._read_status()
    ~~~~~
File "/usr/lib/python3.11/http/client.py", line 286, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
    ~~~~~
File "/usr/lib/python3.11/socket.py", line 718, in readinto
    return self._sock.recv_into(b)
    ~~~~~
File "/usr/lib/python3.11/ssl.py", line 1314, in recv_into
    return self.read(nbytes, buffer)
    ~~~~~
File "/usr/lib/python3.11/ssl.py", line 1166, in read
    return self._sslobj.read(len, buffer)
    ~~~~~
TimeoutError: The read operation timed out
```

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 787, in urlopen
    response = self._make_request(
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 536, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=read_timeout)
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 367, in _raise_timeout
    raise ReadTimeoutError(
urllib3.exceptions.ReadTimeoutError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed
out. (read timeout=1)
```

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

```

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 667, in send
    resp = conn.urlopen(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 841, in urlopen
    retries = retries.increment(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/util/retry.py", line 519, in increment
    raise MaxRetryError(_pool, url, reason) from reason # type: ignore[arg-
type]
    ~~~~~~

urllib3.exceptions.MaxRetryError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

During handling of the above exception, another exception occurred:

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 482, in _request
    resp = self.session.get(url, timeout=timeout, headers=headers)
            ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 602, in get
    return self.request("GET", url, **kwargs)
            ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 589, in request
    resp = self.send(prepare, **send_kwargs)
            ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
            ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-

```

```

packages/requests/adapters.py", line 700, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```

File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 136, in _retries_gen
    yield i # Run the function.
    ~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 274, in __call__
    res = self.func(*args, **kwargs)
    ~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/nominatim.py", line 372, in reverse
    return self._call_geocoder(url, callback, timeout=timeout)
    ~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/base.py", line 368, in _call_geocoder
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)
    ~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 472, in get_json
    resp = self._request(url, timeout=timeout, headers=headers)
    ~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 494, in _request
    raise GeocoderUnavailable(message)
geopy.exc.GeocoderUnavailable:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))
2025-07-07 22:03:13,774 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:16,441 - WARNING - Retrying (Retry(total=1, connect=None,

```

```

read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:17,630 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:18,837 - WARNING - RateLimiter caught an error, retrying (0/2
tries). Called with *((14.39582814, 33.53808821),), **{'exactly_one': True,
'language': 'en'}).
Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 534, in _make_request
    response = conn.getresponse()
                ~~~~~~
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connection.py", line 516, in getresponse
    httplib_response = super().getresponse()
                       ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 1395, in getresponse
    response.begin()
  File "/usr/lib/python3.11/http/client.py", line 325, in begin
    version, status, reason = self._read_status()
                             ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 286, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
           ~~~~~~
  File "/usr/lib/python3.11/socket.py", line 718, in readinto
    return self._sock.recv_into(b)
           ~~~~~~
  File "/usr/lib/python3.11/ssl.py", line 1314, in recv_into
    return self.read(nbytes, buffer)
           ~~~~~~
  File "/usr/lib/python3.11/ssl.py", line 1166, in read
    return self._sslobj.read(len, buffer)
           ~~~~~~
TimeoutError: The read operation timed out

```

The above exception was the direct cause of the following exception:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 787, in urlopen
    response = self._make_request(

```

```

~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 536, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=read_timeout)
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 367, in _raise_timeout
    raise ReadTimeoutError(
urllib3.exceptions.ReadTimeoutError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed
out. (read timeout=1)

```

The above exception was the direct cause of the following exception:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 667, in send
    resp = conn.urlopen(
            ~~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 841, in urlopen
    retries = retries.increment(
            ~~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/util/retry.py", line 519, in increment
    raise MaxRetryError(_pool, url, reason) from reason # type: ignore[arg-
type]
~~~~~
urllib3.exceptions.MaxRetryError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 482, in _request
    resp = self.session.get(url, timeout=timeout, headers=headers)

```

```

~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 602, in get
    return self.request("GET", url, **kwargs)
~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 589, in request
    resp = self.send(prepare, **send_kwargs)
~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 700, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 136, in _retries_gen
    yield i # Run the function.
    ~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 274, in __call__
    res = self.func(*args, **kwargs)
    ~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/nominatim.py", line 372, in reverse
    return self._call_geocoder(url, callback, timeout=timeout)
    ~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/base.py", line 368, in _call_geocoder
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)
    ~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 472, in get_json
    resp = self._request(url, timeout=timeout, headers=headers)
    ~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 494, in _request

```



```

        raise GeocoderUnavailable(message)
geopy.exc.GeocoderUnavailable:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=14.39582814&lon=33.53808821&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))
2025-07-07 22:03:26,847 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=12.90415662&lon=23.47547804&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:28,506 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.46057695&lon=26.12652531&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:29,701 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.46057695&lon=26.12652531&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:31,481 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.46057695&lon=26.12652531&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:33,449 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.46057695&lon=26.12652531&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:35,472 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=14.0387405&lon=35.38913814&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:37,619 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-

```

```

language=en&addressdetails=1
2025-07-07 22:03:39,463 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:40,647 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:42,431 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:43,612 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)")':
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:44,797 - WARNING - RateLimiter caught an error, retrying (0/2
tries). Called with *((15.59742522, 32.53558534),), **{'exactly_one': True,
'language': 'en'}).
Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 534, in _make_request
    response = conn.getresponse()
                ~~~~~~
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connection.py", line 516, in getresponse
    httplib_response = super().getresponse()
                        ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 1395, in getresponse
    response.begin()
  File "/usr/lib/python3.11/http/client.py", line 325, in begin
    version, status, reason = self._read_status()
                                ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 286, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
            ~~~~~~
  File "/usr/lib/python3.11/socket.py", line 718, in readinto
    return self._sock.recv_into(b)
            ~~~~~~

```

```
File "/usr/lib/python3.11/ssl.py", line 1314, in recv_into
    return self.read(nbytes, buffer)
    ~~~~~
```

```
File "/usr/lib/python3.11/ssl.py", line 1166, in read
    return self._sslobj.read(len, buffer)
    ~~~~~
```

TimeoutError: The read operation timed out

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 787, in urlopen
    response = self._make_request(
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 536, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=read_timeout)
```

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 367, in _raise_timeout
    raise ReadTimeoutError(
urllib3.exceptions.ReadTimeoutError:
```

```
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed
out. (read timeout=1)
```

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 667, in send
    resp = conn.urlopen(
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 841, in urlopen
    retries = retries.increment(
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/util/retry.py", line 519, in increment
    raise MaxRetryError(_pool, url, reason) from reason # type: ignore[arg-
type]
```

```

~~~~~
urllib3.exceptions.MaxRetryError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 482, in _request

```

```

    resp = self.session.get(url, timeout=timeout, headers=headers)
    ~~~~~

```

```

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 602, in get
    return self.request("GET", url, **kwargs)
    ~~~~~

```

```

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 589, in request
    resp = self.send(prepare, **send_kwargs)
    ~~~~~

```

```

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
    ~~~~~

```

```

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 700, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 136, in _retries_gen
    yield i # Run the function.
    ~~~~~

```

```

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 274, in __call__
    res = self.func(*args, **kwargs)
    ~~~~~

```

```

File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/nominatim.py", line 372, in reverse
    return self._call_geocoder(url, callback, timeout=timeout)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/base.py", line 368, in _call_geocoder
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 472, in get_json
    resp = self._request(url, timeout=timeout, headers=headers)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 494, in _request
    raise GeocoderUnavailable(message)
geopy.exc.GeocoderUnavailable:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))
2025-07-07 22:03:52,804 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:57,774 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:03:58,968 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:02,673 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:04,603 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',

```

```

port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:05,809 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:06,998 - WARNING - RateLimiter caught an error, retrying (0/2
tries). Called with *((15.59742522, 32.53558534),), **{'exactly_one': True,
'language': 'en'}).
Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 534, in _make_request
    response = conn.getresponse()
                ~~~~~~
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connection.py", line 516, in getresponse
    httplib_response = super().getresponse()
                        ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 1395, in getresponse
    response.begin()
  File "/usr/lib/python3.11/http/client.py", line 325, in begin
    version, status, reason = self._read_status()
                                ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 286, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
            ~~~~~~
  File "/usr/lib/python3.11/socket.py", line 718, in readinto
    return self._sock.recv_into(b)
           ~~~~~~
  File "/usr/lib/python3.11/ssl.py", line 1314, in recv_into
    return self.read(nbytes, buffer)
           ~~~~~~
  File "/usr/lib/python3.11/ssl.py", line 1166, in read
    return self._sslobj.read(len, buffer)
           ~~~~~~
TimeoutError: The read operation timed out

```

The above exception was the direct cause of the following exception:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 787, in urlopen
    response = self._make_request(
                ~~~~~~
  File "/home/hadoop/py311env/lib/python3.11/site-

```

```

packages/urllib3/connectionpool.py", line 536, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=read_timeout)
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 367, in _raise_timeout
    raise ReadTimeoutError(
urllib3.exceptions.ReadTimeoutError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed
out. (read timeout=1)

```

The above exception was the direct cause of the following exception:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 667, in send
    resp = conn.urlopen(
            ~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 841, in urlopen
    retries = retries.increment(
            ~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/util/retry.py", line 519, in increment
    raise MaxRetryError(_pool, url, reason) from reason # type: ignore[arg-
type]
    ~~~~~
urllib3.exceptions.MaxRetryError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 482, in _request
    resp = self.session.get(url, timeout=timeout, headers=headers)
            ~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-

```

```

packages/requests/sessions.py", line 602, in get
    return self.request("GET", url, **kwargs)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 589, in request
    resp = self.send(prepare, **send_kwargs)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 700, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError:
HTTPSPoolConnection(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSPoolConnection(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```

File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 136, in _retries_gen
    yield i # Run the function.
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 274, in __call__
    res = self.func(*args, **kwargs)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/nominatim.py", line 372, in reverse
    return self._call_geocoder(url, callback, timeout=timeout)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/base.py", line 368, in _call_geocoder
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 472, in get_json
    resp = self._request(url, timeout=timeout, headers=headers)
    ~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 494, in _request
    raise GeocoderUnavailable(message)
geopy.exc.GeocoderUnavailable:

```



```
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))
2025-07-07 22:04:15,008 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:16,191 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=15.59742522&lon=32.53558534&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:22,274 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:23,905 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:28,080 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:29,947 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:31,138 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:32,757 - WARNING - Retrying (Retry(total=1, connect=None,
```

```

read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:33,938 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:35,823 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:37,007 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:38,963 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:40,800 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:41,976 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1))':
/reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:04:43,164 - WARNING - RateLimiter caught an error, retrying (0/2
tries). Called with *((13.62471333, 25.35801212),), **{'exactly_one': True,
'language': 'en'}).
Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 534, in _make_request
    response = conn.getresponse()

```

```

~~~~~
File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connection.py", line 516, in getresponse
    httplib_response = super().getresponse()
    ~~~~~
File "/usr/lib/python3.11/http/client.py", line 1395, in getresponse
    response.begin()
File "/usr/lib/python3.11/http/client.py", line 325, in begin
    version, status, reason = self._read_status()
    ~~~~~
File "/usr/lib/python3.11/http/client.py", line 286, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
    ~~~~~
File "/usr/lib/python3.11/socket.py", line 718, in readinto
    return self._sock.recv_into(b)
    ~~~~~
File "/usr/lib/python3.11/ssl.py", line 1314, in recv_into
    return self.read(nbytes, buffer)
    ~~~~~
File "/usr/lib/python3.11/ssl.py", line 1166, in read
    return self._sslobj.read(len, buffer)
    ~~~~~
TimeoutError: The read operation timed out

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 787, in urlopen
    response = self._make_request(
    ~~~~~
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 536, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=read_timeout)
    File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 367, in _raise_timeout
    raise ReadTimeoutError(
urllib3.exceptions.ReadTimeoutError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed
out. (read timeout=1)

```

The above exception was the direct cause of the following exception:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 667, in send
    resp = conn.urlopen(
    ~~~~~

```

```

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
        ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
        ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 841, in urlopen
    retries = retries.increment(
        ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/util/retry.py", line 519, in increment
    raise MaxRetryError(_pool, url, reason) from reason # type: ignore[arg-
type]
    ~~~~~~

urllib3.exceptions.MaxRetryError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=13.62471333&lon=25.35801212&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

During handling of the above exception, another exception occurred:

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 482, in _request
    resp = self.session.get(url, timeout=timeout, headers=headers)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 602, in get
    return self.request("GET", url, **kwargs)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 589, in request
    resp = self.send(prepare, **send_kwargs)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 700, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries

```

```
exceeded with url: /reverse?lat=13.62471333&lon=25.35801212&format=json&accept-  
language=en&addressdetails=1 (Caused by  
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',  
port=443): Read timed out. (read timeout=1)"))
```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-  
packages/geopy/extra/rate_limiter.py", line 136, in _retries_gen  
    yield i # Run the function.  
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-  
packages/geopy/extra/rate_limiter.py", line 274, in __call__  
    res = self.func(*args, **kwargs)  
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-  
packages/geopy/geocoders/nominatim.py", line 372, in reverse  
    return self._call_geocoder(url, callback, timeout=timeout)  
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-  
packages/geopy/geocoders/base.py", line 368, in _call_geocoder  
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)  
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",  
line 472, in get_json  
    resp = self._request(url, timeout=timeout, headers=headers)  
    ~~~~~
```

```
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",  
line 494, in _request
```

```
    raise GeocoderUnavailable(message)
```

geopy.exc.GeocoderUnavailable:

```
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries  
exceeded with url: /reverse?lat=13.62471333&lon=25.35801212&format=json&accept-  
language=en&addressdetails=1 (Caused by  
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',  
port=443): Read timed out. (read timeout=1)"))
```

```
2025-07-07 22:04:55,185 - WARNING - Retrying (Retry(total=1, connect=None,  
read=None, redirect=None, status=None)) after connection broken by  
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',  
port=443): Read timed out. (read timeout=1)")':
```

```
/reverse?lat=13.18240174&lon=30.21860152&format=json&accept-  
language=en&addressdetails=1
```

```
2025-07-07 22:04:56,396 - WARNING - Retrying (Retry(total=0, connect=None,  
read=None, redirect=None, status=None)) after connection broken by  
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',  
port=443): Read timed out. (read timeout=1)")':
```

```
/reverse?lat=13.18240174&lon=30.21860152&format=json&accept-
```

language=en&addressdetails=1
2025-07-07 22:04:58,061 - WARNING - Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=13.18240174&lon=30.21860152&format=json&accept-language=en&addressdetails=1
2025-07-07 22:04:59,245 - WARNING - Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=13.18240174&lon=30.21860152&format=json&accept-language=en&addressdetails=1
2025-07-07 22:05:04,053 - WARNING - Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=17.58736932&lon=33.96306453&format=json&accept-language=en&addressdetails=1
2025-07-07 22:05:05,269 - WARNING - Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=17.58736932&lon=33.96306453&format=json&accept-language=en&addressdetails=1
2025-07-07 22:05:12,121 - WARNING - Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=12.05407646&lon=24.8760273&format=json&accept-language=en&addressdetails=1
2025-07-07 22:05:13,977 - WARNING - Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=12.05407646&lon=24.8760273&format=json&accept-language=en&addressdetails=1
2025-07-07 22:05:16,156 - WARNING - Retrying (Retry(total=1, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=12.05407646&lon=24.8760273&format=json&accept-language=en&addressdetails=1
2025-07-07 22:05:17,337 - WARNING - Retrying (Retry(total=0, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ReadTimeoutError("HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed out. (read timeout=1)")':
/reverse?lat=12.05407646&lon=24.8760273&format=json&accept-

```

language=en&addressdetails=1
2025-07-07 22:05:18,538 - WARNING - RateLimiter caught an error, retrying (0/2
tries). Called with (*((12.05407646, 24.8760273),), **{'exactly_one': True,
'language': 'en'}).
Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 534, in _make_request
    response = conn.getresponse()
                ~~~~~~
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connection.py", line 516, in getresponse
    httplib_response = super().getresponse()
                        ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 1395, in getresponse
    response.begin()
  File "/usr/lib/python3.11/http/client.py", line 325, in begin
    version, status, reason = self._read_status()
                                ~~~~~~
  File "/usr/lib/python3.11/http/client.py", line 286, in _read_status
    line = str(self.fp.readline(_MAXLINE + 1), "iso-8859-1")
            ~~~~~~
  File "/usr/lib/python3.11/socket.py", line 718, in readinto
    return self._sock.recv_into(b)
            ~~~~~~
  File "/usr/lib/python3.11/ssl.py", line 1314, in recv_into
    return self.read(nbytes, buffer)
            ~~~~~~
  File "/usr/lib/python3.11/ssl.py", line 1166, in read
    return self._sslobj.read(len, buffer)
            ~~~~~~
TimeoutError: The read operation timed out

```

The above exception was the direct cause of the following exception:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 787, in urlopen
    response = self._make_request(
                ~~~~~~
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 536, in _make_request
    self._raise_timeout(err=e, url=url, timeout_value=read_timeout)
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 367, in _raise_timeout
    raise ReadTimeoutError(
urllib3.exceptions.ReadTimeoutError:
HTTPConnectionPool(host='nominatim.openstreetmap.org', port=443): Read timed
out. (read timeout=1)

```

The above exception was the direct cause of the following exception:

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 667, in send
    resp = conn.urlopen(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 871, in urlopen
    return self.urlopen(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/connectionpool.py", line 841, in urlopen
    retries = retries.increment(
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/urllib3/util/retry.py", line 519, in increment
    raise MaxRetryError(_pool, url, reason) from reason # type: ignore[arg-
type]
    ~~~~~~

urllib3.exceptions.MaxRetryError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=12.05407646&lon=24.8760273&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))
```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 482, in _request
    resp = self.session.get(url, timeout=timeout, headers=headers)
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 602, in get
    return self.request("GET", url, **kwargs)
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/sessions.py", line 589, in request
    resp = self.send(prepare, **send_kwargs)
            ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
```



```

packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
    ~~~~~~

File "/home/hadoop/py311env/lib/python3.11/site-
packages/requests/adapters.py", line 700, in send
    raise ConnectionError(e, request=request)
requests.exceptions.ConnectionError:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=12.05407646&lon=24.8760273&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))

```

During handling of the above exception, another exception occurred:

```

Traceback (most recent call last):
  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 136, in _retries_gen
    yield i # Run the function.
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/extra/rate_limiter.py", line 274, in __call__
    res = self.func(*args, **kwargs)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/nominatim.py", line 372, in reverse
    return self._call_geocoder(url, callback, timeout=timeout)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-
packages/geopy/geocoders/base.py", line 368, in _call_geocoder
    result = self.adapter.get_json(url, timeout=timeout, headers=req_headers)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 472, in get_json
    resp = self._request(url, timeout=timeout, headers=headers)
    ~~~~~~

  File "/home/hadoop/py311env/lib/python3.11/site-packages/geopy/adapters.py",
line 494, in _request
    raise GeocoderUnavailable(message)
geopy.exc.GeocoderUnavailable:
HTTPSConnectionPool(host='nominatim.openstreetmap.org', port=443): Max retries
exceeded with url: /reverse?lat=12.05407646&lon=24.8760273&format=json&accept-
language=en&addressdetails=1 (Caused by
ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"))
2025-07-07 22:05:24,749 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',

```

port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=12.05407646&lon=24.8760273&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:26,341 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=12.05407646&lon=24.8760273&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:33,520 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=13.43755252&lon=22.44914398&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:35,813 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=13.43755252&lon=22.44914398&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:37,270 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=13.43755252&lon=22.44914398&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:38,461 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=13.43755252&lon=22.44914398&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:40,355 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:41,560 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):':
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:43,470 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',

```

port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:44,653 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:46,730 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:47,929 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:49,560 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:50,745 - WARNING - Retrying (Retry(total=0, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=11.71747777&lon=28.34637592&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:52,603 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.17897728&lon=32.74749408&format=json&accept-
language=en&addressdetails=1
2025-07-07 22:05:55,563 - WARNING - Retrying (Retry(total=1, connect=None,
read=None, redirect=None, status=None)) after connection broken by
'ReadTimeoutError("HTTPSConnectionPool(host='nominatim.openstreetmap.org',
port=443): Read timed out. (read timeout=1)"):
/reverse?lat=13.17897728&lon=32.74749408&format=json&accept-
language=en&addressdetails=1

```



```

# 1. Define class names
class_names = [
    "armed conflict",
    "displacement",
    "looting",
    "sexual violence",
    "infrastructure destruction",
    "protest suppression",
    "famine"
]

def load_and_preprocess_image(entry):
    if entry is None or entry == '' or entry == [] or (isinstance(entry, float)
    ↪and np.isnan(entry)):
        return np.zeros((224, 224, 3), dtype=np.float32)
    if isinstance(entry, str):
        try:
            parsed = ast.literal_eval(entry)
            if isinstance(parsed, list):
                entry = parsed
        except Exception:
            pass
    if isinstance(entry, list):
        if len(entry) == 0:
            return np.zeros((224, 224, 3), dtype=np.float32)
        entry = entry[0]
    if isinstance(entry, str):
        if entry.startswith('http://') or entry.startswith('https://'):
            try:
                response = requests.get(entry, timeout=10)
                img = Image.open(BytesIO(response.content)).convert("RGB").
    ↪resize((224, 224))
                return np.array(img, dtype=np.float32) / 255.0
            except Exception as e:
                print(f"Warning: Could not download image {entry}: {e}")
                return np.zeros((224, 224, 3), dtype=np.float32)
        else:
            try:
                img = Image.open(entry).convert("RGB").resize((224, 224))
                return np.array(img, dtype=np.float32) / 255.0
            except Exception as e:
                print(f"Warning: Could not open image {entry}: {e}")
                return np.zeros((224, 224, 3), dtype=np.float32)
    elif isinstance(entry, np.ndarray):
        if entry.shape == (224, 224, 3):
            return entry.astype(np.float32)

```

```

    try:
        img = Image.fromarray((entry * 255).astype(np.uint8)).convert("RGB")
        img = img.resize((224, 224))
        return np.array(img, dtype=np.float32) / 255.0
    except Exception as e:
        print(f"Warning: Could not process array image: {e}")
        return np.zeros((224, 224, 3), dtype=np.float32)
else:
    print(f"Warning: Unsupported image format: {type(entry)}")
    return np.zeros((224, 224, 3), dtype=np.float32)

# Preprocess images
final_sample_df['image'] = final_sample_df['images'].
    ↪ apply(load_and_preprocess_image)

# Prepare input arrays
texts = final_sample_df['text'].astype(str).values
images = np.stack(final_sample_df['image'].values)

# Model inference
predictions = model.predict([texts, images])
predicted_labels = (predictions > 0.5).astype(int)

# Add class-named columns
for idx, class_name in enumerate(class_names):
    final_sample_df[f'pred_{class_name}'] = predicted_labels[:, idx]
    final_sample_df[f'prob_{class_name}'] = predictions[:, idx]

# Export
final_sample_df.to_csv("threat_detection_predictions.csv", index=False)
print("Exported threat detection predictions to threat_detection_predictions.
    ↪ csv")

```

Warning: Could not open image

"["http://pbs.twimg.com/media/GsfuzYgW8AAQAW6.jpg"]": [Errno 2] No such file or directory: '"["http://pbs.twimg.com/media/GsfuzYgW8AAQAW6.jpg"]"'

3/3 [=====] - 15s 5s/step

Exported threat detection predictions to threat_detection_predictions.csv

[]:

[]: