MF815_RAG_Algo_Zhiqi_Shota

May 3, 2025

Author: Shota Ichitaku

```python
[1]: import numpy as np
     import pandas as pd
```

```python
[2]: !pip install openai
     !pip install langchain
     !pip install langchain-chroma
     !pip install -U langchain-community
     !pip install tiktoken
     !pip install unstructured
     !pip install --upgrade --force-reinstall pandas scipy numba
```

Requirement already satisfied: openai in /usr/local/lib/python3.11/dist-packages
(1.76.0)
Requirement already satisfied: anyio<5,>=3.5.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (4.9.0)
Requirement already satisfied: distro<2,>=1.7.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (1.9.0)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (0.28.1)
Requirement already satisfied: jiter<1,>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (0.9.0)
Requirement already satisfied: pydantic<3,>=1.9.0 in
/usr/local/lib/python3.11/dist-packages (from openai) (2.11.3)
Requirement already satisfied: sniffio in /usr/local/lib/python3.11/dist-
packages (from openai) (1.3.1)
Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.11/dist-packages
(from openai) (4.67.1)
Requirement already satisfied: typing-extensions<5,>=4.11 in
/usr/local/lib/python3.11/dist-packages (from openai) (4.13.2)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.11/dist-
packages (from anyio<5,>=3.5.0->openai) (3.10)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-
packages (from httpx<1,>=0.23.0->openai) (2025.1.31)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx<1,>=0.23.0->openai) (1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-

packages (from httpcore==1.*->httpx<1,>=0.23.0->openai) (0.16.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3,>=1.9.0->openai)
(0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3,>=1.9.0->openai)
(2.33.1)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3,>=1.9.0->openai)
(0.4.0)
Requirement already satisfied: langchain in /usr/local/lib/python3.11/dist-
packages (0.3.24)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.55 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.56)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.8 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.8)
Requirement already satisfied: langsmith<0.4,>=0.1.17 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.34)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in
/usr/local/lib/python3.11/dist-packages (from langchain) (2.11.3)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.11/dist-packages (from langchain) (2.0.40)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-
packages (from langchain) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-
packages (from langchain) (6.0.2)
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.55->langchain) (9.1.2)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.55->langchain) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.55->langchain) (24.2)
Requirement already satisfied: typing-extensions>=4.7 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.55->langchain) (4.13.2)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(0.28.1)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(3.10.16)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in

/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(0.23.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain)
(0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain)
(2.33.1)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain)
(0.4.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests<3,>=2->langchain) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain)
(2025.1.31)
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-
packages (from SQLAlchemy<3,>=1.4->langchain) (3.2.1)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain) (4.9.0)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain) (1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from
httpcore==1.*->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain) (0.16.0)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core<1.0.0,>=0.3.55->langchain) (3.0.0)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain)
(1.3.1)
Collecting langchain-chroma
  Downloading langchain_chroma-0.2.3-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: langchain-core>=0.3.52 in
/usr/local/lib/python3.11/dist-packages (from langchain-chroma) (0.3.56)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.11/dist-
packages (from langchain-chroma) (2.0.2)
Collecting chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0
.7.0,>=0.4.0 (from langchain-chroma)
  Downloading chromadb-0.6.3-py3-none-any.whl.metadata (6.8 kB)
Collecting build>=1.0.3 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5
,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading build-1.2.2.post1-py3-none-any.whl.metadata (6.5 kB)
Requirement already satisfied: pydantic>=1.9 in /usr/local/lib/python3.11/dist-

packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (2.11.3)
Collecting chroma-hnswlib==0.7.6 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading chroma_hnswlib-0.7.6-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (252 bytes)
Collecting fastapi>=0.95.2 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading fastapi-0.115.12-py3-none-any.whl.metadata (27 kB)
Collecting uvicorn>=0.18.3 (from uvicorn[standard]>=0.18.3->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading uvicorn-0.34.2-py3-none-any.whl.metadata (6.5 kB)
Collecting posthog>=2.4.0 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading posthog-4.0.0-py2.py3-none-any.whl.metadata (3.0 kB)
Requirement already satisfied: typing_extensions>=4.5.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (4.13.2)
Collecting onnxruntime>=1.14.1 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading onnxruntime-1.21.1-cp311-cp311-manylinux_2_27_x86_64.manylinux_2_28_x86_64.whl.metadata (4.5 kB)
Requirement already satisfied: opentelemetry-api>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.32.1)
Collecting opentelemetry-exporter-otlp-proto-grpc>=1.2.0 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading opentelemetry_exporter_otlp_proto_grpc-1.32.1-py3-none-any.whl.metadata (2.5 kB)
Collecting opentelemetry-instrumentation-fastapi>=0.41b0 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading opentelemetry_instrumentation_fastapi-0.53b1-py3-none-any.whl.metadata (2.2 kB)
Requirement already satisfied: opentelemetry-sdk>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.32.1)
Requirement already satisfied: tokenizers>=0.13.2 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.21.1)
Collecting pypika>=0.48.9 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading PyPika-0.48.9.tar.gz (67 kB)
                              67.3/67.3 kB
5.6 MB/s eta 0:00:00
  Installing build dependencies … done

```
  Getting requirements to build wheel … done
  Preparing metadata (pyproject.toml) … done
Requirement already satisfied: tqdm>=4.65.0 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (4.67.1)
Collecting overrides>=7.3.1 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0
.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading overrides-7.7.0-py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: importlib-resources in
/usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12
,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (6.5.2)
Requirement already satisfied: grpcio>=1.58.0 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (1.71.0)
Collecting bcrypt>=4.0.1 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.
5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading bcrypt-4.3.0-cp39-abi3-manylinux_2_34_x86_64.whl.metadata (10 kB)
Requirement already satisfied: typer>=0.9.0 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (0.15.2)
Collecting kubernetes>=28.1.0 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!
=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading kubernetes-32.0.1-py2.py3-none-any.whl.metadata (1.7 kB)
Requirement already satisfied: tenacity>=8.2.3 in
/usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12
,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (9.1.2)
Requirement already satisfied: PyYAML>=6.0.0 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (6.0.2)
Collecting mmh3>=4.0.1 (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,
!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma)
  Downloading mmh3-5.1.0-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
4.whl.metadata (16 kB)
Requirement already satisfied: orjson>=3.9.12 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (3.10.16)
Requirement already satisfied: httpx>=0.27.0 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (0.28.1)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.11/dist-
packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.
9,<0.7.0,>=0.4.0->langchain-chroma) (13.9.4)
Requirement already satisfied: langsmith<0.4,>=0.1.125 in
/usr/local/lib/python3.11/dist-packages (from langchain-core>=0.3.52->langchain-
chroma) (0.3.34)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-core>=0.3.52->langchain-
```

chroma) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-core>=0.3.52->langchain-
chroma) (24.2)
Collecting pyproject_hooks (from build>=1.0.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading pyproject_hooks-1.2.0-py3-none-any.whl.metadata (1.3 kB)
Collecting starlette<0.47.0,>=0.40.0 (from fastapi>=0.95.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading starlette-0.46.2-py3-none-any.whl.metadata (6.2 kB)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (4.9.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-
packages (from httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2025.1.31)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.0.9)
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages
(from httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.10)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from httpcore==1.*->httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.16.0)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core>=0.3.52->langchain-chroma) (3.0.0)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.11/dist-
packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.17.0)
Requirement already satisfied: python-dateutil>=2.5.3 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.9.0.post0)
Requirement already satisfied: google-auth>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.38.0)
Requirement already satisfied: websocket-

client!=0.40.0,!=0.41.*,!=0.42.*,>=0.32.0 in /usr/local/lib/python3.11/dist-
packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.8.0)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-
packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.32.3)
Requirement already satisfied: requests-oauthlib in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.0.0)
Requirement already satisfied: oauthlib>=3.2.2 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.2.2)
Requirement already satisfied: urllib3>=1.24.2 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.4.0)
Collecting durationpy>=0.7 (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading durationpy-0.9-py3-none-any.whl.metadata (338 bytes)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-core>=0.3.52->langchain-chroma) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-core>=0.3.52->langchain-chroma) (0.23.0)
Collecting coloredlogs (from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading coloredlogs-15.0.1-py2.py3-none-any.whl.metadata (12 kB)
Requirement already satisfied: flatbuffers in /usr/local/lib/python3.11/dist-
packages (from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (25.2.10)
Requirement already satisfied: protobuf in /usr/local/lib/python3.11/dist-
packages (from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (5.29.4)
Requirement already satisfied: sympy in /usr/local/lib/python3.11/dist-packages
(from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.13.1)
Requirement already satisfied: deprecated>=1.2.6 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-api>=1.2.0-

```
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.2.18)
Requirement already satisfied: importlib-metadata<8.7.0,>=6.0 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-api>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (8.6.1)
Requirement already satisfied: googleapis-common-protos~=1.52 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.70.0)
Collecting opentelemetry-exporter-otlp-proto-common==1.32.1 (from opentelemetry-
exporter-otlp-proto-grpc>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading opentelemetry_exporter_otlp_proto_common-1.32.1-py3-none-
any.whl.metadata (1.9 kB)
Collecting opentelemetry-proto==1.32.1 (from opentelemetry-exporter-otlp-proto-g
rpc>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading opentelemetry_proto-1.32.1-py3-none-any.whl.metadata (2.4 kB)
Collecting opentelemetry-instrumentation-asgi==0.53b1 (from opentelemetry-
instrumentation-fastapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading opentelemetry_instrumentation_asgi-0.53b1-py3-none-
any.whl.metadata (2.1 kB)
Collecting opentelemetry-instrumentation==0.53b1 (from opentelemetry-
instrumentation-fastapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading opentelemetry_instrumentation-0.53b1-py3-none-any.whl.metadata
(6.8 kB)
Requirement already satisfied: opentelemetry-semantic-conventions==0.53b1 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-instrumentation-fast
api>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.53b1)
Collecting opentelemetry-util-http==0.53b1 (from opentelemetry-instrumentation-f
astapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading opentelemetry_util_http-0.53b1-py3-none-any.whl.metadata (2.6 kB)
Requirement already satisfied: wrapt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-
instrumentation==0.53b1->opentelemetry-instrumentation-fastapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
```

.0->langchain-chroma) (1.17.2)
Collecting asgiref~=3.0 (from opentelemetry-instrumentation-
asgi==0.53b1->opentelemetry-instrumentation-fastapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading asgiref-3.8.1-py3-none-any.whl.metadata (9.3 kB)
Collecting monotonic>=1.5 (from posthog>=2.4.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading monotonic-1.6-py2.py3-none-any.whl.metadata (1.5 kB)
Collecting backoff>=1.10.0 (from posthog>=2.4.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading backoff-2.2.1-py3-none-any.whl.metadata (14 kB)
Requirement already satisfied: distro>=1.5.0 in /usr/local/lib/python3.11/dist-
packages (from posthog>=2.4.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.9.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=1.9-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=1.9-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.33.1)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=1.9-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.4.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.11/dist-packages (from rich>=10.11.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/dist-packages (from rich>=10.11.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.19.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in
/usr/local/lib/python3.11/dist-packages (from tokenizers>=0.13.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.30.2)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.11/dist-
packages (from typer>=0.9.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (8.1.8)
Requirement already satisfied: shellingham>=1.3.0 in
/usr/local/lib/python3.11/dist-packages (from typer>=0.9.0-

>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.5.4)
Collecting httptools>=0.6.3 (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading httptools-0.6.4-cp311-cp311-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
4.whl.metadata (3.6 kB)
Collecting python-dotenv>=0.13 (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading python_dotenv-1.1.0-py3-none-any.whl.metadata (24 kB)
Collecting uvloop!=0.15.0,!=0.15.1,>=0.14.0 (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading uvloop-0.21.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.9 kB)
Collecting watchfiles>=0.13 (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading watchfiles-1.0.5-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.9 kB)
Requirement already satisfied: websockets>=10.4 in
/usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (15.0.1)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from google-auth>=1.0.1-
>kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (5.5.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.11/dist-packages (from google-auth>=1.0.1-
>kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.4.2)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-
packages (from google-auth>=1.0.1->kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (4.9.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-
packages (from huggingface-hub<1.0,>=0.16.4->tokenizers>=0.13.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.18.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4-
>tokenizers>=0.13.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4

```
.0->langchain-chroma) (2025.3.2)
Requirement already satisfied: zipp>=3.20 in /usr/local/lib/python3.11/dist-
packages (from importlib-metadata<8.7.0,>=6.0->opentelemetry-api>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.21.0)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-
packages (from markdown-it-py>=2.2.0->rich>=10.11.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.4.1)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.3.1)
Collecting humanfriendly>=9.1 (from coloredlogs->onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma)
  Downloading humanfriendly-10.0-py2.py3-none-any.whl.metadata (9.2 kB)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy->onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.3.0)
Requirement already satisfied: pyasn1<0.7.0,>=0.6.1 in
/usr/local/lib/python3.11/dist-packages (from pyasn1-modules>=0.2.1->google-auth
>=1.0.1->kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.6.1)
Downloading langchain_chroma-0.2.3-py3-none-any.whl (11 kB)
Downloading chromadb-0.6.3-py3-none-any.whl (611 kB)
                        611.1/611.1 kB
14.6 MB/s eta 0:00:00
Downloading
chroma_hnswlib-0.7.6-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(2.4 MB)
                        2.4/2.4 MB
54.0 MB/s eta 0:00:00
Downloading bcrypt-4.3.0-cp39-abi3-manylinux_2_34_x86_64.whl (284 kB)
                        284.2/284.2 kB
23.9 MB/s eta 0:00:00
Downloading build-1.2.2.post1-py3-none-any.whl (22 kB)
Downloading fastapi-0.115.12-py3-none-any.whl (95 kB)
                        95.2/95.2 kB
8.7 MB/s eta 0:00:00
Downloading kubernetes-32.0.1-py2.py3-none-any.whl (2.0 MB)
                        2.0/2.0 MB
```

```
watchfiles-1.0.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (454
kB)
                              454.8/454.8 kB
34.9 MB/s eta 0:00:00
Downloading coloredlogs-15.0.1-py2.py3-none-any.whl (46 kB)
                              46.0/46.0 kB
4.4 MB/s eta 0:00:00
Downloading pyproject_hooks-1.2.0-py3-none-any.whl (10 kB)
Downloading asgiref-3.8.1-py3-none-any.whl (23 kB)
Downloading humanfriendly-10.0-py2.py3-none-any.whl (86 kB)
                              86.8/86.8 kB
6.8 MB/s eta 0:00:00
Building wheels for collected packages: pypika
  Building wheel for pypika (pyproject.toml) … done
  Created wheel for pypika: filename=pypika-0.48.9-py2.py3-none-any.whl
size=53800
sha256=e2a7c44c908308348e29cf943be08b652447e919dad39157078bb5adba628e7c
  Stored in directory: /root/.cache/pip/wheels/a3/01/bd/4c40ceb9d5354160cb186dcc
153360f4ab7eb23e2b24daf96d
Successfully built pypika
Installing collected packages: pypika, monotonic, durationpy, uvloop, uvicorn,
python-dotenv, pyproject_hooks, overrides, opentelemetry-util-http,
opentelemetry-proto, mmh3, humanfriendly, httptools, chroma-hnswlib, bcrypt,
backoff, asgiref, watchfiles, starlette, posthog, opentelemetry-exporter-otlp-
proto-common, coloredlogs, build, onnxruntime, kubernetes, fastapi,
opentelemetry-instrumentation, opentelemetry-instrumentation-asgi,
opentelemetry-exporter-otlp-proto-grpc, opentelemetry-instrumentation-fastapi,
chromadb, langchain-chroma
Successfully installed asgiref-3.8.1 backoff-2.2.1 bcrypt-4.3.0
build-1.2.2.post1 chroma-hnswlib-0.7.6 chromadb-0.6.3 coloredlogs-15.0.1
durationpy-0.9 fastapi-0.115.12 httptools-0.6.4 humanfriendly-10.0
kubernetes-32.0.1 langchain-chroma-0.2.3 mmh3-5.1.0 monotonic-1.6
onnxruntime-1.21.1 opentelemetry-exporter-otlp-proto-common-1.32.1
opentelemetry-exporter-otlp-proto-grpc-1.32.1 opentelemetry-
instrumentation-0.53b1 opentelemetry-instrumentation-asgi-0.53b1 opentelemetry-
instrumentation-fastapi-0.53b1 opentelemetry-proto-1.32.1 opentelemetry-util-
http-0.53b1 overrides-7.7.0 posthog-4.0.0 pypika-0.48.9 pyproject_hooks-1.2.0
python-dotenv-1.1.0 starlette-0.46.2 uvicorn-0.34.2 uvloop-0.21.0
watchfiles-1.0.5
Collecting langchain-community
  Downloading langchain_community-0.3.23-py3-none-any.whl.metadata (2.5 kB)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.56 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.56)
Requirement already satisfied: langchain<1.0.0,>=0.3.24 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.24)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (2.0.40)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-
```

packages (from langchain-community) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (3.11.15)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (9.1.2)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain-community)
  Downloading dataclasses_json-0.6.7-py3-none-any.whl.metadata (25 kB)
Collecting pydantic-settings<3.0.0,>=2.4.0 (from langchain-community)
  Downloading pydantic_settings-2.9.1-py3-none-any.whl.metadata (3.8 kB)
Requirement already satisfied: langsmith<0.4,>=0.1.125 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.34)
Collecting httpx-sse<1.0.0,>=0.4.0 (from langchain-community)
  Downloading httpx_sse-0.4.0-py3-none-any.whl.metadata (9.0 kB)
Requirement already satisfied: numpy>=1.26.2 in /usr/local/lib/python3.11/dist-packages (from langchain-community) (2.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (6.4.3)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.20.0)
Collecting marshmallow<4.0.0,>=3.18.0 (from dataclasses-json<0.7,>=0.5.7->langchain-community)
  Downloading marshmallow-3.26.1-py3-none-any.whl.metadata (7.3 kB)
Collecting typing-inspect<1,>=0.4.0 (from dataclasses-json<0.7,>=0.5.7->langchain-community)
  Downloading typing_inspect-0.9.0-py3-none-any.whl.metadata (1.5 kB)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.8 in /usr/local/lib/python3.11/dist-packages (from langchain<1.0.0,>=0.3.24->langchain-community) (0.3.8)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in /usr/local/lib/python3.11/dist-packages (from

langchain<1.0.0,>=0.3.24->langchain-community) (2.11.3)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.56->langchain-community) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.56->langchain-community) (24.2)
Requirement already satisfied: typing-extensions>=4.7 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.56->langchain-community) (4.13.2)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (0.28.1)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (3.10.16)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (0.23.0)
Requirement already satisfied: python-dotenv>=0.21.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic-
settings<3.0.0,>=2.4.0->langchain-community) (1.1.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic-
settings<3.0.0,>=2.4.0->langchain-community) (0.4.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-
community) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests<3,>=2->langchain-community) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-
community) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-
community) (2025.1.31)
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-
packages (from SQLAlchemy<3,>=1.4->langchain-community) (3.2.1)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community) (4.9.0)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community)
(1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from

httpcore==1.*->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community)
(0.16.0)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core<1.0.0,>=0.3.56->langchain-community) (3.0.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from
pydantic<3.0.0,>=2.7.4->langchain<1.0.0,>=0.3.24->langchain-community) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from
pydantic<3.0.0,>=2.7.4->langchain<1.0.0,>=0.3.24->langchain-community) (2.33.1)
Collecting mypy-extensions>=0.3.0 (from typing-inspect<1,>=0.4.0->dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Downloading mypy_extensions-1.1.0-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-
community) (1.3.1)
Downloading langchain_community-0.3.23-py3-none-any.whl (2.5 MB)
                        2.5/2.5 MB
31.4 MB/s eta 0:00:00
Downloading dataclasses_json-0.6.7-py3-none-any.whl (28 kB)
Downloading httpx_sse-0.4.0-py3-none-any.whl (7.8 kB)
Downloading pydantic_settings-2.9.1-py3-none-any.whl (44 kB)
                        44.4/44.4 kB
3.7 MB/s eta 0:00:00
Downloading marshmallow-3.26.1-py3-none-any.whl (50 kB)
                        50.9/50.9 kB
4.1 MB/s eta 0:00:00
Downloading typing_inspect-0.9.0-py3-none-any.whl (8.8 kB)
Downloading mypy_extensions-1.1.0-py3-none-any.whl (5.0 kB)
Installing collected packages: mypy-extensions, marshmallow, httpx-sse, typing-
inspect, pydantic-settings, dataclasses-json, langchain-community
Successfully installed dataclasses-json-0.6.7 httpx-sse-0.4.0 langchain-
community-0.3.23 marshmallow-3.26.1 mypy-extensions-1.1.0 pydantic-
settings-2.9.1 typing-inspect-0.9.0
Collecting tiktoken
  Downloading tiktoken-0.9.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.7 kB)
Requirement already satisfied: regex>=2022.1.18 in
/usr/local/lib/python3.11/dist-packages (from tiktoken) (2024.11.6)
Requirement already satisfied: requests>=2.26.0 in
/usr/local/lib/python3.11/dist-packages (from tiktoken) (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests>=2.26.0->tiktoken) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in

/usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken)
(2025.1.31)
Downloading
tiktoken-0.9.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2
MB)
                          1.2/1.2 MB
14.3 MB/s eta 0:00:00
Installing collected packages: tiktoken
Successfully installed tiktoken-0.9.0
Collecting unstructured
  Downloading unstructured-0.17.2-py3-none-any.whl.metadata (24 kB)
Requirement already satisfied: chardet in /usr/local/lib/python3.11/dist-
packages (from unstructured) (5.2.0)
Collecting filetype (from unstructured)
  Downloading filetype-1.2.0-py2.py3-none-any.whl.metadata (6.5 kB)
Collecting python-magic (from unstructured)
  Downloading python_magic-0.4.27-py2.py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: lxml in /usr/local/lib/python3.11/dist-packages
(from unstructured) (5.4.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages
(from unstructured) (3.9.1)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-
packages (from unstructured) (2.32.3)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.11/dist-
packages (from unstructured) (4.13.4)
Collecting emoji (from unstructured)
  Downloading emoji-2.14.1-py3-none-any.whl.metadata (5.7 kB)
Requirement already satisfied: dataclasses-json in
/usr/local/lib/python3.11/dist-packages (from unstructured) (0.6.7)
Collecting python-iso639 (from unstructured)
  Downloading python_iso639-2025.2.18-py3-none-any.whl.metadata (14 kB)
Collecting langdetect (from unstructured)
  Downloading langdetect-1.0.9.tar.gz (981 kB)
                          981.5/981.5 kB
16.0 MB/s eta 0:00:00
  Preparing metadata (setup.py) … done
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages
(from unstructured) (2.0.2)
Collecting rapidfuzz (from unstructured)
  Downloading rapidfuzz-3.13.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Requirement already satisfied: backoff in /usr/local/lib/python3.11/dist-
packages (from unstructured) (2.2.1)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.11/dist-packages (from unstructured) (4.13.2)

```
Collecting unstructured-client (from unstructured)
  Downloading unstructured_client-0.34.0-py3-none-any.whl.metadata (21 kB)
Requirement already satisfied: wrapt in /usr/local/lib/python3.11/dist-packages
(from unstructured) (1.17.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages
(from unstructured) (4.67.1)
Requirement already satisfied: psutil in /usr/local/lib/python3.11/dist-packages
(from unstructured) (5.9.5)
Collecting python-oxmsg (from unstructured)
  Downloading python_oxmsg-0.0.2-py3-none-any.whl.metadata (5.0 kB)
Requirement already satisfied: html5lib in /usr/local/lib/python3.11/dist-
packages (from unstructured) (1.1)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.11/dist-
packages (from beautifulsoup4->unstructured) (2.7)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in
/usr/local/lib/python3.11/dist-packages (from dataclasses-json->unstructured)
(3.26.1)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from dataclasses-json->unstructured)
(0.9.0)
Requirement already satisfied: six>=1.9 in /usr/local/lib/python3.11/dist-
packages (from html5lib->unstructured) (1.17.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.11/dist-
packages (from html5lib->unstructured) (0.5.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages
(from nltk->unstructured) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages
(from nltk->unstructured) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.11/dist-packages (from nltk->unstructured) (2024.11.6)
Collecting olefile (from python-oxmsg->unstructured)
  Downloading olefile-0.47-py2.py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->unstructured) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests->unstructured) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->unstructured) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->unstructured)
(2025.1.31)
Collecting aiofiles>=24.1.0 (from unstructured-client->unstructured)
  Downloading aiofiles-24.1.0-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: cryptography>=3.1 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(43.0.3)
Collecting eval-type-backport>=0.2.0 (from unstructured-client->unstructured)
  Downloading eval_type_backport-0.2.2-py3-none-any.whl.metadata (2.2 kB)
```

```
Requirement already satisfied: httpx>=0.27.0 in /usr/local/lib/python3.11/dist-
packages (from unstructured-client->unstructured) (0.28.1)
Requirement already satisfied: nest-asyncio>=1.6.0 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(1.6.0)
Requirement already satisfied: pydantic>=2.11.2 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(2.11.3)
Collecting pypdf>=4.0 (from unstructured-client->unstructured)
  Downloading pypdf-5.4.0-py3-none-any.whl.metadata (7.3 kB)
Requirement already satisfied: requests-toolbelt>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(1.0.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(0.4.0)
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.11/dist-
packages (from cryptography>=3.1->unstructured-client->unstructured) (1.17.1)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx>=0.27.0->unstructured-client->unstructured) (4.9.0)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx>=0.27.0->unstructured-client->unstructured) (1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from httpcore==1.*->httpx>=0.27.0->unstructured-client->unstructured)
(0.16.0)
Requirement already satisfied: packaging>=17.0 in
/usr/local/lib/python3.11/dist-packages (from
marshmallow<4.0.0,>=3.18.0->dataclasses-json->unstructured) (24.2)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=2.11.2->unstructured-
client->unstructured) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=2.11.2->unstructured-
client->unstructured) (2.33.1)
Requirement already satisfied: mypy-extensions>=0.3.0 in
/usr/local/lib/python3.11/dist-packages (from typing-
inspect<1,>=0.4.0->dataclasses-json->unstructured) (1.1.0)
Requirement already satisfied: pycparser in /usr/local/lib/python3.11/dist-
packages (from cffi>=1.12->cryptography>=3.1->unstructured-client->unstructured)
(2.22)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx>=0.27.0->unstructured-client->unstructured) (1.3.1)
Downloading unstructured-0.17.2-py3-none-any.whl (1.8 MB)
                        1.8/1.8 MB
61.0 MB/s eta 0:00:00
Downloading emoji-2.14.1-py3-none-any.whl (590 kB)
                        590.6/590.6 kB
44.1 MB/s eta 0:00:00
```

```
Downloading filetype-1.2.0-py2.py3-none-any.whl (19 kB)
Downloading python_iso639-2025.2.18-py3-none-any.whl (167 kB)
                        167.6/167.6 kB
15.8 MB/s eta 0:00:00
Downloading python_magic-0.4.27-py2.py3-none-any.whl (13 kB)
Downloading python_oxmsg-0.0.2-py3-none-any.whl (31 kB)
Downloading
rapidfuzz-3.13.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.1
MB)
                        3.1/3.1 MB
83.8 MB/s eta 0:00:00
Downloading unstructured_client-0.34.0-py3-none-any.whl (189 kB)
                        189.4/189.4 kB
17.6 MB/s eta 0:00:00
Downloading aiofiles-24.1.0-py3-none-any.whl (15 kB)
Downloading eval_type_backport-0.2.2-py3-none-any.whl (5.8 kB)
Downloading pypdf-5.4.0-py3-none-any.whl (302 kB)
                        302.3/302.3 kB
27.8 MB/s eta 0:00:00
Downloading olefile-0.47-py2.py3-none-any.whl (114 kB)
                        114.6/114.6 kB
11.0 MB/s eta 0:00:00
Building wheels for collected packages: langdetect
  Building wheel for langdetect (setup.py) … done
  Created wheel for langdetect: filename=langdetect-1.0.9-py3-none-any.whl
size=993223
sha256=df8d8acbc06b82a9bfdaa29229831d45c0aeb699e3ad5bb7f0f7f1ff6ac582d8
  Stored in directory: /root/.cache/pip/wheels/0a/f2/b2/e5ca405801e05eb7c8ed5b3b
4bcf1fcabcd6272c167640072e
Successfully built langdetect
Installing collected packages: filetype, rapidfuzz, python-magic, python-iso639,
pypdf, olefile, langdetect, eval-type-backport, emoji, aiofiles, python-oxmsg,
unstructured-client, unstructured
Successfully installed aiofiles-24.1.0 emoji-2.14.1 eval-type-backport-0.2.2
filetype-1.2.0 langdetect-1.0.9 olefile-0.47 pypdf-5.4.0 python-iso639-2025.2.18
python-magic-0.4.27 python-oxmsg-0.0.2 rapidfuzz-3.13.0 unstructured-0.17.2
unstructured-client-0.34.0
Collecting pandas
  Downloading
pandas-2.2.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(89 kB)
                        89.9/89.9 kB
2.6 MB/s eta 0:00:00
Collecting scipy
  Downloading
scipy-1.15.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(61 kB)
                        62.0/62.0 kB
```

```
4.8 MB/s eta 0:00:00
Collecting numba
  Downloading
numba-0.61.2-cp311-cp311-manylinux2014_x86_64.manylinux_2_17_x86_64.whl.metadata
(2.8 kB)
Collecting numpy>=1.23.2 (from pandas)
  Downloading
numpy-2.2.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata
(62 kB)
                          62.0/62.0 kB
5.7 MB/s eta 0:00:00
Collecting python-dateutil>=2.8.2 (from pandas)
  Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz>=2020.1 (from pandas)
  Downloading pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
  Downloading tzdata-2025.2-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting llvmlite<0.45,>=0.44.0dev0 (from numba)
  Downloading llvmlite-0.44.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.8 kB)
Collecting six>=1.5 (from python-dateutil>=2.8.2->pandas)
  Downloading six-1.17.0-py2.py3-none-any.whl.metadata (1.7 kB)
Downloading
pandas-2.2.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (13.1
MB)
                          13.1/13.1 MB
82.5 MB/s eta 0:00:00
Downloading
scipy-1.15.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (37.6
MB)
                          37.6/37.6 MB
24.0 MB/s eta 0:00:00
Downloading
numba-0.61.2-cp311-cp311-manylinux2014_x86_64.manylinux_2_17_x86_64.whl (3.8 MB)
                          3.8/3.8 MB
88.1 MB/s eta 0:00:00
Downloading
llvmlite-0.44.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (42.4
MB)
                          42.4/42.4 MB
16.8 MB/s eta 0:00:00
Downloading
numpy-2.2.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (16.4 MB)
                          16.4/16.4 MB
96.4 MB/s eta 0:00:00
Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
                          229.9/229.9 kB
16.8 MB/s eta 0:00:00
```

```
Downloading pytz-2025.2-py2.py3-none-any.whl (509 kB)
                              509.2/509.2 kB
34.2 MB/s eta 0:00:00
Downloading tzdata-2025.2-py2.py3-none-any.whl (347 kB)
                              347.8/347.8 kB
24.8 MB/s eta 0:00:00
Downloading six-1.17.0-py2.py3-none-any.whl (11 kB)
Installing collected packages: pytz, tzdata, six, numpy, llvmlite, scipy,
python-dateutil, numba, pandas
  Attempting uninstall: pytz
    Found existing installation: pytz 2025.2
    Uninstalling pytz-2025.2:
      Successfully uninstalled pytz-2025.2
  Attempting uninstall: tzdata
    Found existing installation: tzdata 2025.2
    Uninstalling tzdata-2025.2:
      Successfully uninstalled tzdata-2025.2
  Attempting uninstall: six
    Found existing installation: six 1.17.0
    Uninstalling six-1.17.0:
      Successfully uninstalled six-1.17.0
  Attempting uninstall: numpy
    Found existing installation: numpy 2.0.2
    Uninstalling numpy-2.0.2:
      Successfully uninstalled numpy-2.0.2
  Attempting uninstall: llvmlite
    Found existing installation: llvmlite 0.43.0
    Uninstalling llvmlite-0.43.0:
      Successfully uninstalled llvmlite-0.43.0
  Attempting uninstall: scipy
    Found existing installation: scipy 1.15.2
    Uninstalling scipy-1.15.2:
      Successfully uninstalled scipy-1.15.2
  Attempting uninstall: python-dateutil
    Found existing installation: python-dateutil 2.9.0.post0
    Uninstalling python-dateutil-2.9.0.post0:
      Successfully uninstalled python-dateutil-2.9.0.post0
  Attempting uninstall: numba
    Found existing installation: numba 0.60.0
    Uninstalling numba-0.60.0:
      Successfully uninstalled numba-0.60.0
  Attempting uninstall: pandas
    Found existing installation: pandas 2.2.2
    Uninstalling pandas-2.2.2:
      Successfully uninstalled pandas-2.2.2
```

Successfully installed llvmlite-0.44.0 numba-0.61.2 numpy-2.2.5 pandas-2.2.3
python-dateutil-2.9.0.post0 pytz-2025.2 scipy-1.15.2 six-1.17.0 tzdata-2025.2

```
[3]: import os
     import openai
     import sys
     sys.path.append('../..')

     os.environ['OPENAI_API_KEY'] =  ''
```

```
[4]: # Uncomment if you aare using colab. Now we connect the colab notebook to our
     ↪drive to access the data.
     from google.colab import drive
     drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

```
[5]: %cd /content/gdrive/MyDrive/'Colab Notebooks'/MF815/NLP/'NLP_app '
```

/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app

```
[6]: !ls
```

```
df_RAG2_5.csv              MutualFundSummary2
df_RAG2.csv                MutualFundSummary2_5
```

```
df_RAG3_5.csv                MutualFundSummary3
df_RAG3.csv                  MutualFundSummary3_5
df_RAG4.csv                  MutualFundSummary4
df_RAG_all.csv               MutualFundSummary4_5
df_RAG.csv                   MutualFundSummary5
docs                         MutualFundSummary6
FundArticlesHealthy.csv      MutualFundSummary7
FundArticlesUnhealthy.csv    MutualFundSummary8
MutualFundLabels.csv         MutualFundSummary9
MutualFundSummary            NLP_Deep_Learning_Questrom_WS_03_2020.ipynb
MutualFundSummary1           word2vec_perso.txt
MutualFundSummary10
```

[7]:
```python
from IPython.display import HTML
```

[8]:
```python
# Progress bar
def progress(value, max=100):
    return HTML("""
        <progress
            value='{value}'
            max='{max}',
            style='width: 100%'
        >
            {value}
        </progress>
    """.format(value=value, max=max))


# read the repo in PATH and append the texts in a list
def get_data(PATH):
    list_dir = os.listdir(PATH)
    texts = []
    fund_names = []
    out = display(progress(0, len(list_dir)-1), display_id=True)
    for ii, filename in enumerate(list_dir) :
        with open(PATH+'/'+filename, 'r', encoding="utf8") as f :
            txt = f.read()
            try :
                txt_split = txt.split('<head_breaker>')
                summary = txt_split[1].strip()
                fund_name = txt_split[0].strip()
            except :
                summary = txt
                fund_name = ''
        texts.append(summary)
        fund_names.append(fund_name)
        out.update(progress(ii, len(list_dir)-1))
    return fund_names, texts
```

```
[9]:  # Get the summaries

      fund_names, summaries = get_data('MutualFundSummary')
```

<IPython.core.display.HTML object>

```
[10]: #fund_names
```

```
[68]: print(len(fund_names))
```

545

```
[11]: df_extraction = pd.DataFrame({'fund_name' : fund_names})
      df_label = pd.read_csv('MutualFundLabels.csv')
      df = df_label.merge(df_extraction, on='fund_name', how='inner')
      df_label = df_label.rename(columns={'Leverage?': 'Derivatives_label'})
      df_label
```

```
[11]:                      id                                   fund_name   \
      0    0000051931-18-000151              American Funds College 2018 Fund
      1    0000051931-18-000151              American Funds College 2021 Fund
      2    0000051931-18-000151              American Funds College 2024 Fund
      3    0000051931-18-000151              American Funds College 2027 Fund
      4    0000051931-18-000151              American Funds College 2030 Fund
      ..                    …                                           …
      462  0001710607-18-000172      American Century Diversified Corporate Bond ETF
      463  0001710607-18-000172      American Century Diversified Municipal Bond ETF
      464  0001710607-18-000172    American Century Quality Diversified Internati…
      465  0001710607-18-000172       American Century STOXX U.S. Quality Growth ETF
      466  0001710607-18-000172        American Century STOXX U.S. Quality Value ETF

          Performance fee?              Ivestment Strategy Derivatives_label   \
      0                NaN         Balanced Fund (Low Risk)             Yes
      1                NaN         Balanced Fund (Low Risk)             Yes
      2                NaN         Balanced Fund (Low Risk)             Yes
      3                NaN         Balanced Fund (Low Risk)             Yes
      4                NaN         Balanced Fund (Low Risk)             Yes
      ..                …                              …                …
      462              NaN  Fixed Income Long Only (Low Risk)             Yes
      463              NaN  Fixed Income Long Only (Low Risk)              No
      464              NaN       Equity Long Only (Low Risk)             Yes
      465              NaN       Equity Long Only (Low Risk)             Yes
      466              NaN       Equity Long Only (Low Risk)             Yes

                      Portfolio composition  Concentration
      0           Investment grade securities     Diversified
      1           Investment grade securities     Diversified
```

```
2                        Investment grade securities    Diversified
3                        Investment grade securities    Diversified
4                        Investment grade securities    Diversified
..                                               …            …
462                      Investment grade securities    Diversified
463                      Investment grade securities    Diversified
464  Sub-investment grade securities or emerging ma…    Diversified
465                                  Listed Equities    Diversified
466                                  Listed Equities    Diversified

[467 rows x 7 columns]
```

[12]: `df`

[12]:
```
                      id                                       fund_name  \
0    0000051931-18-000151                American Funds College 2018 Fund
1    0000051931-18-000151                American Funds College 2021 Fund
2    0000051931-18-000151                American Funds College 2024 Fund
3    0000051931-18-000151                American Funds College 2027 Fund
4    0000051931-18-000151                American Funds College 2030 Fund
..                    …                                                …
461  0001710607-18-000172     American Century Diversified Corporate Bond ETF
462  0001710607-18-000172     American Century Diversified Municipal Bond ETF
463  0001710607-18-000172   American Century Quality Diversified Internati…
464  0001710607-18-000172      American Century STOXX U.S. Quality Growth ETF
465  0001710607-18-000172       American Century STOXX U.S. Quality Value ETF

    Performance fee?                Ivestment Strategy Leverage?  \
0                NaN          Balanced Fund (Low Risk)        Yes
1                NaN          Balanced Fund (Low Risk)        Yes
2                NaN          Balanced Fund (Low Risk)        Yes
3                NaN          Balanced Fund (Low Risk)        Yes
4                NaN          Balanced Fund (Low Risk)        Yes
..                 …                                 …          …
461              NaN  Fixed Income Long Only (Low Risk)        Yes
462              NaN  Fixed Income Long Only (Low Risk)         No
463              NaN        Equity Long Only (Low Risk)        Yes
464              NaN        Equity Long Only (Low Risk)        Yes
465              NaN        Equity Long Only (Low Risk)        Yes

                      Portfolio composition  Concentration
0             Investment grade securities    Diversified
1             Investment grade securities    Diversified
2             Investment grade securities    Diversified
3             Investment grade securities    Diversified
4             Investment grade securities    Diversified
..                                      …              …
```

```
461                        Investment grade securities    Diversified
462                        Investment grade securities    Diversified
463  Sub-investment grade securities or emerging ma…    Diversified
464                                   Listed Equities    Diversified
465                                   Listed Equities    Diversified

[466 rows x 7 columns]
```

0.1 Ask what investment strategy do they take

```python
[70]: from langchain.document_loaders import DirectoryLoader

      # Load documents from a directory
      loader = DirectoryLoader(path="./MutualFundSummary", glob="*.txt")

      # Load the text data from the directory
      documents = loader.load()
```

```python
[71]: #print(len(documents))
```

```python
[72]: #document = documents[0]
```

```python
[73]: #print(document)
```

```python
[78]: from langchain.text_splitter import RecursiveCharacterTextSplitter,␣
      ↪CharacterTextSplitter
      from langchain_chroma import Chroma
      from langchain.vectorstores import Chroma
      from langchain.embeddings.openai import OpenAIEmbeddings
      embedding = OpenAIEmbeddings()
      persist_directory = './docs/chroma'

      #llm_name = "gpt-3.5-turbo"
      llm_name = "gpt-4o-mini"
      print(llm_name)

      from langchain.chat_models import ChatOpenAI
      llm = ChatOpenAI(model_name=llm_name, temperature=0)
      from langchain.chains import RetrievalQA
```

```
gpt-4o-mini
```

```python
[79]: r_splitter = RecursiveCharacterTextSplitter(
          chunk_size=1000,
          chunk_overlap=100,
          separators=["\n\n", "\n", "(?<=\. )", " "]
      )
```

```python
[80]:  #splits=r_splitter.split_documents(documents)
       ii = 4

       splits=r_splitter.split_documents([documents[ii]])

       fund_name = fund_names[ii]

       persist_directory = f"./docs/chroma/{fund_name}"

       vectordb = Chroma.from_documents(
         documents=splits,
         embedding=embedding,
         persist_directory=persist_directory
       )

       vectordb.persist()

       print(len(vectordb))
```

       341

```python
[82]:  from langchain.prompts import PromptTemplate
```

```python
[83]:  # Build prompt
       template = """Use the following pieces of context to answer the question at the
        ↪end. Please answer only in following terms: Balanced Fund (LowRisk) or Fixed
        ↪Income Long Only (Low Risk) or
       Equity Long Only (Low Risk) or Long Short Funds (High Risk) or Commodities Fund
        ↪(Low Risk) . If you don't know the answer, just say that you don't know,
        ↪don't try to make up an answer.
       {context}
       Question: {question}
       Helpful Answer:"""
       QA_CHAIN_PROMPT = PromptTemplate.from_template(template)
```

```python
[85]:  qa_chain = RetrievalQA.from_chain_type(
           llm,
           retriever=vectordb.as_retriever(),
           return_source_documents=True,
           chain_type_kwargs={"prompt": QA_CHAIN_PROMPT}
       )
```

```python
[86]:  #fund_name
```

```python
[87]:  #question = f"What category of fund can {fund_name} be classified as? "
       #question = f"Who are the portfolio managers of Ultra-Short Bond Fund?"
```

```
#result = qa_chain({"query": question})
```

[88]:
```
#result
```

[89]:
```
#print(result["source_documents"][0].metadata)
```

[90]:
```
df_RAG = pd.DataFrame(fund_names, columns=['fund_name'])
df_RAG['Predicted investment strategy'] = np.nan
df_RAG['investment strategy evidence'] = np.nan
df_RAG['source meta data'] = np.nan
```

[91]:
```
df_RAG
```

[91]:
```
                                              fund_name  \
0                                   AB Arizona Portfolio
1                         AB Asia ex-Japan Equity Portfolio
2                             AB Credit Long/Short Portfolio
3                        AB Balanced Wealth Strategy Portfolio
4                                 AB Bond Inflation Strategy
..                                                     …
540  Variable Portfolio - Victory Sycamore Established Value Fund
541                 Variable Portfolio - TCW Core Plus Bond Fund
542  Western Asset Intermediate Maturity New York Municipals Fund
543                 Western Asset Massachusetts Municipals Fund
544                                          World Growth Fund

    Predicted investment strategy  investment strategy evidence  \
0                             NaN                           NaN
1                             NaN                           NaN
2                             NaN                           NaN
3                             NaN                           NaN
4                             NaN                           NaN
..                            …                             …
540                           NaN                           NaN
541                           NaN                           NaN
542                           NaN                           NaN
543                           NaN                           NaN
544                           NaN                           NaN

    source meta data
0                NaN
1                NaN
2                NaN
3                NaN
4                NaN
..               …
540              NaN
```

```
541              NaN
542              NaN
543              NaN
544              NaN

[545 rows x 4 columns]
```

```
[43]: import time
```

```
[92]: for ii in range(len(df_RAG)):
        #print(ii)

        fund_name = df_RAG.iloc[ii]['fund_name']

        print(fund_name)

        splits=r_splitter.split_documents([documents[ii]])

        persist_directory = f"./docs/chroma/{fund_name}"

        vectordb = Chroma.from_documents(
          documents=splits,
          embedding=embedding,
          persist_directory=persist_directory
        )

        vectordb.persist()

        print(len(vectordb))

        # Build prompt
        template = """Use the following pieces of context to answer the question at␣
        ↪the end. Please answer only in following terms: Balanced Fund (LowRisk) or␣
        ↪Fixed Income Long Only (Low Risk) or
Equity Long Only (Low Risk) or Long Short Funds (High Risk) or Commodities Fund␣
        ↪(Low Risk) . If you don't know the answer, just say that you don't know,␣
        ↪don't try to make up an answer.
{context}
Question: {question}
Helpful Answer:"""
        QA_CHAIN_PROMPT = PromptTemplate.from_template(template)

        qa_chain = RetrievalQA.from_chain_type(
          llm,
          retriever=vectordb.as_retriever(),
          return_source_documents=True,
          chain_type_kwargs={"prompt": QA_CHAIN_PROMPT}
```

```
    )

    question = f"What category of fund can {fund_name} be classified as? "


    result = qa_chain({"query": question})

    #print(result["result"])

    df_RAG.iloc[ii, 1] = result["result"]

    df_RAG.iloc[ii, 2] = result["source_documents"][0].page_content

    df_RAG.iloc[ii, 3] = str(result["source_documents"][0].metadata)

    # --- ADD THE DELAY HERE ---
        # This pause happens *after* all processing for index 'ii' is done,
        # including the API call, before starting the next iteration (ii+1).
    #print(f"--- Finished processing index {ii}. Waiting for 21 seconds before␣
    ↪next API call... ---")
    #time.sleep(21) # Pause for 10 seconds
        # ------------------------
```

AB Arizona Portfolio
297

<ipython-input-92-0973c680d478>:44: FutureWarning: Setting an item of
incompatible dtype is deprecated and will raise an error in a future version of
pandas. Value 'Fixed Income Long Only (Low Risk)' has dtype incompatible with
float64, please explicitly cast to a compatible dtype first.
  df_RAG.iloc[ii, 1] = result["result"]
<ipython-input-92-0973c680d478>:46: FutureWarning: Setting an item of
incompatible dtype is deprecated and will raise an error in a future version of
pandas. Value 'AB Arizona Portfolio

<head_breaker>

AB Arizona Portfolio

INVESTMENT OBJECTIVE: The investment objective of the Portfolio is to earn the
highest level of current income exempt from both federal income tax and State of
Arizona personal income tax that is available without assuming what the Adviser
considers to be undue risk.

FEES AND EXPENSES OF THE PORTFOLIO:' has dtype incompatible with float64, please
explicitly cast to a compatible dtype first.

```
  df_RAG.iloc[ii, 2] = result["source_documents"][0].page_content
<ipython-input-92-0973c680d478>:48: FutureWarning: Setting an item of
incompatible dtype is deprecated and will raise an error in a future version of
pandas. Value '{'source': 'MutualFundSummary/AB Arizona Portfolio.txt'}' has
dtype incompatible with float64, please explicitly cast to a compatible dtype
first.
  df_RAG.iloc[ii, 3] = str(result["source_documents"][0].metadata)
```

AB Asia ex-Japan Equity Portfolio
279
AB Credit Long/Short Portfolio
403
AB Balanced Wealth Strategy Portfolio
270
AB Bond Inflation Strategy
372
AB All Market Income Portfolio
203
AB All Market Real Return Portfolio
266
AB Discovery Value Fund
162
AB Dynamic Asset Allocation Portfolio
135
AB FlexFee International Strategic Core Portfolio
104
AB FlexFee Core Opportunities Portfolio
120
AB FlexFee Emerging Markets Growth Portfolio
85
AB Global Risk Allocation-Moderate Portfolio
104
AB FlexFee Large Cap Growth Portfolio
96
AB Global Thematic Growth Portfolio
68
AB Income Fund
124
AB Intermediate Bond Portfolio
128
AB Limited Duration High Income Portfolio
120
AB Massachusetts Portfolio
144
AB Minnesota Portfolio
132
AB Municipal Bond Inflation Strategy
116

AB Tax-Aware Fixed Income Portfolio

124

AB New Jersey Portfolio

136

AB New York Portfolio

140

AB Real Estate Investment Portfolio

136

AB Pennsylvania Portfolio

136

AB Small/Mid Cap Value Portfolio

76

AB Small Cap Value Portfolio

92

ADAPTIVE EQUITY FUND

64

AC ALTERNATIVES EQUITY MARKET NEUTRAL FUND

80

AB Virginia Portfolio

140

AC ALTERNATIVES DISCIPLINED LONG SHORT FUND

80

Aggressive Growth Fund

68

AllianzGI Emerging Markets Opportunities Fund

96

ALL CAP GROWTH FUND

76

AllianzGI Health Sciences Fund

100

AllianzGI Focused Growth Fund

100

AllianzGI Mid-Cap Fund

92

AllianzGI Global Natural Resources Fund

108

AllianzGI NFJ Dividend Value Fund

96

AllianzGI Global Small-Cap Fund

116

AllianzGI Income & Growth Fund

100

AllianzGI NFJ International Value Fund

96

AllianzGI NFJ Large-Cap Value Fund

96

AllianzGI NFJ Mid-Cap Value Fund

104

American Century Diversified Municipal Bond ETF
76
AllianzGI NFJ Small-Cap Value Fund
100
American Century Quality Diversified International ETF
112
American Century STOXX U.S. Quality Growth ETF
84
AllianzGI Small-Cap Fund
96
American Century Diversified Corporate Bond ETF
68
AllianzGI Technology Fund
96
American Century Growth ETF
48
American Funds 2015 Target Date Retirement Fund
172
American Century Value ETF
48
American Funds 2045 Target Date Retirement Fund
172
American Century STOXX U.S. Quality Value ETF
96
American Funds 2010 Target Date Retirement Fund
172
American Funds 2020 Target Date Retirement Fund
172
American Funds 2030 Target Date Retirement Fund
172
American Funds 2025 Target Date Retirement Fund
172
American Funds 2035 Target Date Retirement Fund
172
American Funds 2040 Target Date Retirement Fund
172
American Funds 2050 Target Date Retirement Fund
172
American Funds 2055 Target Date Retirement Fund
172
American Funds 2060 Target Date Retirement Fund
172
American Funds College 2018 Fund
144
American Funds College 2021 Fund
160
American Funds College Enrollment Fund
128

American Funds College 2024 Fund
156
American Funds College 2027 Fund
144
American Funds College 2033 Fund
144
American Funds College 2030 Fund
144
American Funds Conservative Growth and Income Portfolio
176
American Funds College 2036 Fund
136
American Funds Retirement Income Portfolio - Conservative
176
American Funds Preservation Portfolio
168
American Funds Moderate Growth and Income Portfolio
188
American Funds Retirement Income Portfolio - Enhanced
176
Anchor Tactical Equity Strategies Fund
96
American Funds Tax-Advantaged Growth and Income Portfolio
164
American Funds Retirement Income Portfolio - Moderate
180
American Funds Tax-Exempt Preservation Portfolio
136
Asset Allocation Fund
88
American Funds U.S. Government Money Market Fund
88
Anchor Tactical Credit Strategies Fund
108
BALANCED FUND
96
Anchor Tactical Municipal Strategies Fund
120
BBH Core Select
64
BBH Partner Fund - International Equity
96
Capital Growth Fund
60
BBH Global Core Select
92
Bond Fund
124

BBH Intermediate Municipal Bond Fund
112
BBH Limited Duration Fund
88
Blue Chip Income and Growth Fund
68
CALIFORNIA TAX-FREE MONEY MARKET FUND
64
CALIFORNIA HIGH-YIELD MUNICIPAL FUND
76
CALIFORNIA INTERMEDIATE-TERM TAX-FREE BOND FUND
72
Capital Income Builder
72
CAPITAL VALUE FUND
64
ClearBridge Global Health Care Innovations Fund
428
Columbia Dividend Opportunity Fund
300
Columbia Commodity Strategy Fund
434
Columbia Flexible Capital Income Fund
318
Columbia Large Cap Value Fund
519
Columbia Large Cap Enhanced Core Fund
286
Columbia High Yield Bond Fund
277
Columbia Large Cap Index Fund
222
Columbia Large Cap Growth Fund III
201
Columbia Overseas Value Fund
255
Columbia Mid Cap Index Fund
164
Columbia Mid Cap Value Fund
165
Columbia Mortgage Opportunities Fund
287
Columbia Select Global Growth Fund
183
Columbia Select International Equity Fund
207
Columbia Select Large Cap Equity Fund
169

Columbia Small/Mid Cap Value Fund
186
Columbia Seligman Communications and Information Fund
174
Columbia Small Cap Index Fund
162
Columbia Variable Portfolio – Balanced Fund
195
Columbia Small Cap Value Fund II
158
Columbia Variable Portfolio – Disciplined Core Fund
124
Columbia Variable Portfolio – Dividend Opportunity Fund
157
Columbia Variable Portfolio – Emerging Markets Fund
170
Columbia Variable Portfolio – Global Bond Fund
170
Columbia Variable Portfolio – Mid Cap Growth Fund
90
Columbia Variable Portfolio – Large Cap Index Fund
114
Columbia Variable Portfolio – Mid Cap Value Fund
90
Columbia Variable Portfolio – Select Large-Cap Value Fund
106
Columbia Variable Portfolio – High Yield Bond Fund
116
Columbia Variable Portfolio – Government Money Market Fund
76
Columbia Variable Portfolio – Income Opportunities Fund
116
Columbia Variable Portfolio – Intermediate Bond Fund
140
Columbia Variable Portfolio – Large Cap Growth Fund
80
Columbia Variable Portfolio – U.S. Government Mortgage Fund
128
Columbia Variable Portfolio – Select Smaller-Cap Value Fund
98
CORE PLUS FUND
88
CORE EQUITY PLUS FUND
88
Cornerstone Moderately Aggressive Fund
106
Cornerstone Moderately Conservative Fund
106

DELAWARE EMERGING MARKETS FUND
102
Corporate Bond Fund
100
DELAWARE GLOBAL VALUE FUND
82
Delaware International Small Cap Fund
90
DELAWARE INTERNATIONAL VALUE EQUITY FUND
128
DISCIPLINED GROWTH FUND
86
DIVERSIFIED BOND FUND
122
Dreyfus Alternative Diversifier Strategies Fund
189
Eaton Vance Connecticut Municipal Income Fund
58
Dreyfus Global Equity Income Fund
82
Dreyfus Global Emerging Markets Fund
168
Dreyfus International Bond Fund
63
Eaton Vance Dividend Builder Fund
60
Eaton Vance Global Small-Cap Fund
103
Eaton Vance Growth Fund
70
Dreyfus Select Managers Long/Short Fund
60
Dreyfus Yield Enhancement Strategy Fund
118
Eaton Vance Large-Cap Value Fund
58
Eaton Vance Arizona Municipal Income Fund
115
Eaton Vance Minnesota Municipal Income Fund
72
Eaton Vance TABS 5-to-15 Year Laddered Municipal Bond Fund
32
Eaton Vance Real Estate Fund
26
Eaton Vance Pennsylvania Municipal Income Fund
37
Eaton Vance Small-Cap Fund
26

Eaton Vance New Jersey Municipal Income Fund
72
Eaton Vance TABS 1-to-10 Year Laddered Municipal Bond Fund
84
Eaton Vance TABS 10-to-20 Year Laddered Municipal Bond Fund
30
Eaton Vance Municipal Opportunities Fund
35
Eaton Vance Special Equities Fund
52
Eaton Vance TABS Intermediate-Term Municipal Bond Fund
29
Emerging Markets Fund
80
EQUITY GROWTH FUND
30
Eaton Vance TABS Short-Term Municipal Bond Fund
58
Federated MDT All Cap Core Fund
48
Federated MDT Balanced Fund
80
EMERGING MARKETS SMALL CAP FUND
19
Federated MDT Large Cap Growth Fund
26
Federated MDT Small Cap Core Fund
26
Fidelity Advisor Communications Equipment Fund
30
Federated MDT Small Cap Growth Fund
54
Fidelity Advisor Consumer Discretionary Fund
15
Fidelity Advisor Energy Fund
15
Fidelity Advisor Freedom 2005 Fund
36
Fidelity Advisor Financial Services Fund
28
Fidelity Advisor Freedom 2010 Fund
18
Fidelity Advisor Freedom 2015 Fund
19
Fidelity Advisor Freedom 2020 Fund
19
Fidelity Advisor Freedom 2055 Fund
18

Fidelity Advisor Freedom 2045 Fund
36
Fidelity Advisor Freedom 2050 Fund
18
Fidelity Advisor Freedom 2040 Fund
36
Fidelity Advisor Freedom 2035 Fund
36
Fidelity Advisor Freedom 2025 Fund
38
Fidelity Advisor Freedom 2030 Fund
19
Fidelity Advisor Freedom 2060 Fund
36
Fidelity Advisor Health Care Fund
14
Fidelity Advisor Semiconductors Fund
15
Fidelity Advisor Utilities Fund
30
Fidelity Advisor Industrials Fund
30
Fidelity Advisor New Insights Fund
16
Fidelity Advisor Series Opportunistic Insights Fund
18
Fidelity Advisor Technology Fund
32
Fidelity Advisor Series Equity Value Fund
11
Fidelity Contrafund K6
24
Fidelity Contrafund
18
Fidelity Diversified International Fund
20
Fidelity Global Equity Income Fund
24
Fidelity Emerging Europe, Middle East, Africa (EMEA) Fund
38
Fidelity Emerging Asia Fund
24
Fidelity Emerging Markets Discovery Fund
24
Fidelity Flex Opportunistic Insights Fund
18
Fidelity Flex Short-Term Bond Fund
22

Fidelity Emerging Markets Fund
24
Fidelity Flex U.S. Bond Index Fund
22
Fidelity Flex Conservative Income Bond Fund
22
Fidelity GNMA Fund
24
Fidelity International Discovery Fund
30
Fidelity Investment Grade Bond Fund
40
Fidelity International Capital Appreciation Fund
22
Fidelity Managed Retirement 2005 Fund
42
Fidelity International Growth Fund
32
Fidelity Managed Retirement 2015 Fund
44
Fidelity Managed Retirement 2020 Fund
44
Fidelity Managed Retirement 2025 Fund
44
Fidelity Managed Retirement 2010 Fund
44
Fidelity Managed Retirement Income Fund
40
Fidelity Municipal Income 2019 Fund
28
Fidelity Municipal Income 2023 Fund
30
Fidelity Municipal Income 2021 Fund
30
Fidelity Municipal Income 2025 Fund
24
Fidelity Simplicity RMD 2015 Fund
44
Fidelity Series Opportunistic Insights Fund
18
Fidelity Series Short-Term Credit Fund
12
Fidelity Simplicity RMD 2010 Fund
44
Fidelity Series Corporate Bond Fund
20
Fidelity Simplicity RMD 2005 Fund
21

Fidelity Simplicity RMD 2020 Fund
22
Fidelity ZERO Extended Market Index Fund
10
Fidelity Simplicity RMD Income Fund
20
Fidelity ZERO Large Cap Index Fund
10
FOCUSED INTERNATIONAL GROWTH FUND
16
Franklin FTSE Asia ex Japan ETF
30
Franklin FTSE Australia ETF
29
Franklin FTSE Brazil ETF
31
Franklin FTSE China ETF
31
Franklin FTSE Canada ETF
28
Franklin FTSE Europe ETF
26
Franklin FTSE Germany ETF
27
Franklin FTSE Europe Hedged ETF
36
Franklin FTSE France ETF
27
Franklin FTSE Hong Kong ETF
28
Franklin FTSE Switzerland ETF
28
Franklin FTSE India ETF
29
Franklin FTSE South Korea ETF
29
Franklin FTSE Taiwan ETF
29
Franklin FTSE Italy ETF
29
Franklin FTSE United Kingdom ETF
27
Franklin FTSE Mexico ETF
29
Franklin FTSE Japan Hedged ETF
38
Franklin FTSE Japan ETF
28

Franklin FTSE Russia ETF
31
Franklin Liberty Investment Grade Corporate ETF
28
Franklin Liberty Intermediate Municipal Opportunities ETF
24
Franklin Liberty International Opportunities ETF
29
Franklin Liberty Municipal Bond ETF
23
Franklin Liberty U.S. Low Volatility ETF
17
Franklin LibertyQ Emerging Markets ETF
32
Franklin LibertyQ Global Equity ETF
33
Franklin LibertyQ Global Dividend ETF
34
Franklin LibertyQ International Equity Hedged ETF
40
Franklin Payout 2019 Fund
19
Franklin LibertyQ U.S. Equity ETF
23
Franklin LibertyQ U.S. Mid Cap Equity ETF
23
Franklin LibertyQ U.S. Small Cap Equity ETF
23
Franklin Payout 2018 Fund
20
GLOBAL GOLD FUND
16
Franklin Payout 2020 Fund
19
Franklin Payout 2021 Fund
19
Global Growth Fund
36
Global Growth and Income Fund
20
GLOBAL SMALL CAP FUND
17
Franklin Payout 2022 Fund
18
Global Bond Fund
31
Franklin Payout 2023 Fund
16

Global Balanced Fund

26

Global Small Capitalization Fund

35

Government Securities Fund

23

Growth & Income Fund

18

Growth and Tax Strategy Fund

21

Income Stock Fund

22

Growth-Income Fund

15

Growth Fund

35

INCOME & GROWTH FUND

15

High Income Fund

27

High-Income Bond Fund

22

HERITAGE FUND

18

Intermediate-Term Bond Fund

23

International Fund

17

INTERNATIONAL DISCOVERY FUND

17

INTERNATIONAL CORE EQUITY FUND

16

International Growth and Income Fund

19

INTERNATIONAL OPPORTUNITIES FUND

18

INTERNATIONAL VALUE FUND

19

Ivy European Opportunities Fund

31

INTERNATIONAL GROWTH FUND

17

Ivy Asset Strategy Fund

59

Ivy Core Equity Fund

33

Ivy Balanced Fund

37

Ivy Emerging Markets Equity Fund
48
Ivy Energy Fund
33
Ivy Global Equity Income Fund
36
Ivy Managed International Opportunities Fund
34
Ivy Mid Cap Income Opportunities Fund
29
Ivy Global Growth Fund
37
Ivy Large Cap Growth Fund
33
Ivy Micro Cap Growth Fund
30
Ivy Global Income Allocation Fund
54
Ivy Mid Cap Growth Fund
29
Ivy High Income Fund
47
Ivy LaSalle Global Risk-Managed Real Estate Fund
44
Ivy Limited-Term Bond Fund
32
Ivy Municipal Bond Fund
38
Ivy Municipal High Income Fund
41
Ivy Science and Technology Fund
40
Ivy Natural Resources Fund
44
Ivy Securian Core Bond Fund
39
Ivy Securian Real Estate Securities Fund
36
Ivy Small Cap Core Fund
33
Ivy Small Cap Growth Fund
35
Janus Henderson Emerging Markets Fund
35
Janus Henderson Balanced Fund
40
Ivy Tax-Managed Equity Fund
35

Janus Henderson Contrarian Fund
29
Ivy Value Fund
30
Janus Henderson Asia Equity Fund
36
Janus Henderson Balanced Portfolio
40
Janus Henderson Enterprise Fund
31
Janus Henderson European Focus Fund
42
Janus Henderson Forty Fund
42
Janus Henderson Flexible Bond Portfolio
27
Janus Henderson Enterprise Portfolio
31
Janus Henderson Global Equity Income Fund
44
Janus Henderson Forty Portfolio
18
Janus Henderson Global Bond Portfolio
33
Janus Henderson Global Research Fund
36
Janus Henderson Global Real Estate Fund
36
Janus Henderson Global Life Sciences Fund
36
Janus Henderson Global Research Portfolio
25
Janus Henderson Global Select Fund
38
Janus Henderson Growth and Income Fund
33
Janus Henderson Global Technology Fund
36
Janus Henderson Global Technology Portfolio
27
Janus Henderson Global Unconstrained Bond Portfolio
41
Janus Henderson International Opportunities Fund
44
Janus Henderson International Value Fund
29
Janus Henderson Mid Cap Value Portfolio
19

Janus Henderson Global Value Fund
37
Janus Henderson International Small Cap Fund
38
Janus Henderson Venture Fund
30
Janus Henderson Overseas Fund
38
JPMorgan Equity Focus Fund
31
Janus Henderson Overseas Portfolio
27
JPMorgan Access Balanced Fund
50
Janus Henderson Triton Fund
31
Janus Henderson Research Portfolio
22
Janus Henderson Research Fund
30
Janus Henderson U.S. Growth Opportunities Fund
30
JPMorgan California Tax Free Bond Fund
30
JPMorgan Europe Dynamic Fund
27
JPMorgan Global Allocation Fund
47
MainStay VP Epoch U.S. Small Cap Portfolio
21
JPMorgan New York Tax Free Bond Fund
31
MainStay VP Balanced Portfolio
42
JPMorgan Tax Aware Equity Fund
26
MainStay VP Common Stock Portfolio
23
MainStay VP Eagle Small Growth Portfolio
21
MainStay VP Convertible Portfolio
30
MainStay VP Emerging Markets Equity Portfolio
39
MainStay VP Epoch U.S. Equity Yield Portfolio
24
MainStay VP Growth Allocation Portfolio
52

MainStay VP Income Builder Portfolio
52
MainStay VP International Equity Portfolio
22
MainStay VP Janus Henderson Balanced Portfolio
30
MainStay VP Large Cap Growth Portfolio
21
MainStay VP Mid Cap Core Portfolio
24
MainStay VP MFS Utilities Portfolio
37
MainStay VP S&P 500 Index Portfolio
22
Managed Allocation Fund
19
MainStay VP T. Rowe Price Equity Income Portfolio
30
MainStay VP Small Cap Core Portfolio
21
Managed Risk Asset Allocation Fund
37
Managed Risk Blue Chip Income and Growth Fund
37
Managed Risk Growth Fund
35
Managed Risk Growth-Income Fund
36
Managed Risk International Fund
37
MFS Arkansas Municipal Bond Fund
23
MFS Alabama Municipal Bond Fund
24
MFS Global Equity Fund
17
MFS California Municipal Bond Fund
24
MFS Georgia Municipal Bond Fund
24
MFS Massachusetts Municipal Bond Fund
22
MFS Global Growth Fund
20
MFS Maryland Municipal Bond Fund
24
MFS Global Total Return Fund
26

MFS Municipal Income Fund

21

MFS Strategic Income Fund

24

MFS Mississippi Municipal Bond Fund

23

MFS Utilities Fund

23

MFS Virginia Municipal Bond Fund

24

MFS Tennessee Municipal Bond Fund

24

MFS North Carolina Municipal Bond Fund

22

MFS New York Municipal Bond Fund

23

MFS Pennsylvania Municipal Bond Fund

22

MFS South Carolina Municipal Bond Fund

23

MFS West Virginia Municipal Bond Fund

23

MULTI-ASSET REAL RETURN FUND

27

Money Market Fund

13

NORTHERN U.S. GOVERNMENT SELECT MONEY MARKET FUND

23

Mortgage Fund

27

Multi-Manager Value Strategies Fund

46

Nasdaq-100 Index Fund

19

New World Fund

41

NT CORE EQUITY PLUS FUND

13

NORTHERN U.S. GOVERNMENT MONEY MARKET FUND

22

NORTHERN MONEY MARKET FUND

36

NORTHERN MUNICIPAL MONEY MARKET FUND

37

NT GROWTH FUND

14

NT EQUITY GROWTH FUND

10

NT EMERGING MARKETS FUND

14

NT HERITAGE FUND

14

NT INTERNATIONAL VALUE FUND

15

NT INTERNATIONAL SMALL-MID CAP FUND

13

NT SMALL COMPANY

13

PIMCO California Municipal Bond Fund

26

PIMCO California Short Duration Municipal Income Fund

24

PIMCO California Intermediate Municipal Bond Fund

25

PIMCO Gurtin National Municipal Opportunistic Value Fund

24

PIMCO Gurtin California Municipal Intermediate Value Fund

24

PIMCO Gurtin California Municipal Opportunistic Value Fund

24

PIMCO Gurtin National Municipal Intermediate Value Fund

24

PIMCO REALPATH 2020 Fund

44

PIMCO REALPATH 2025 Fund

43

PIMCO REALPATH 2050 Fund

44

PIMCO REALPATH 2055

44

PIMCO REALPATH 2030 Fund

44

PIMCO REALPATH 2035 Fund

43

Pioneer Dynamic Credit Fund

75

PIMCO REALPATH 2045 Fund

43

PIMCO REALPATH 2040 Fund

44

Pioneer Fundamental Growth Fund

44

Pioneer Global High Yield Fund

63

Portfolio Series - American Funds Managed Risk Growth Portfolio

42

Pioneer Global Multisector Income Fund
70
S&P 500 Index Fund
17
Precious Metals and Minerals Fund
19
PRIME MONEY MARKET FUND
14
Pioneer Multi-Asset Ultrashort Income Fund
73
Portfolio Series - American Funds Growth and Income Portfolio
25
Portfolio Series - American Funds Global Growth Portfolio
27
Portfolio Series - American Funds Managed Risk Growth and Income Portfolio
39
Portfolio Series - American Funds Managed Risk Global Allocation Portfolio
43
Schwab California Municipal Money Fund
27
Schwab New York Municipal Money Fund
27
SELECT FUND
17
Science & Technology Fund
20
SHORT DURATION INFLATION PROTECTION BOND FUND
21
Short-Term Bond Fund
24
SMALL CAP GROWTH FUND
17
Small Cap Stock Fund
18
Swan Defined Risk Emerging Markets Fund
28
SUSTAINABLE EQUITY FUND
15
Swan Defined Risk Foreign Developed Fund
26
Swan Defined Risk Fund
24
T. Rowe Price Africa & Middle East Fund
32
Swan Defined Risk U.S. Small Cap Fund
22
T. Rowe Price Institutional Emerging Markets Equity Fund
21

T. Rowe Price Asia Opportunities Fund
34
T. Rowe Price Emerging Europe Fund
30
T. Rowe Price Emerging Markets Stock Fund
34
T. Rowe Price Institutional Africa & Middle East Fund
25
T. Rowe Price Institutional Global Focused Growth Equity Fund
19
T. Rowe Price Emerging Markets Value Stock Fund
30
T. Rowe Price Institutional Frontier Markets Equity Fund
25
T. Rowe Price Institutional Emerging Markets Bond Fund
20
T. Rowe Price Institutional Global Growth Equity Fund
20
T. Rowe Price Institutional Global Value Equity Fund
17
Ultra-Short Bond Fund
20
T. Rowe Price Institutional International Growth Equity Fund
18
T. Rowe Price Institutional International Bond Fund
22
Thrivent Partner Healthcare Portfolio
22
Total Return Strategy Fund
25
T. Rowe Price Institutional International Concentrated Equity Fund
18
T. Rowe Price Institutional International Core Equity Fund
17
U.S. Government/AAA-Rated Securities Fund
26
Treasury Money Market Trust
13
ULTRA FUND
17
USAA MSCI Emerging Markets Value Momentum Blend Index ETF
34
USAA MSCI International Value Momentum Blend Index ETF
27
USAA MSCI USA Small Cap Value Momentum Blend Index ETF
27
USAA MSCI USA Value Momentum Blend Index ETF
27

UTILITIES FUND

14

Vanguard Intermediate-Term Treasury Fund

11

Vanguard GNMA Fund

13

Vanguard Health Care Fund

14

Vanguard Energy Fund

14

Vanguard Intermediate-Term Investment-Grade Fund

13

Vanguard Intermediate-Term Corporate Bond Index Fund

13

Vanguard Long-Term Corporate Bond Index Fund

13

Vanguard Long-Term Investment-Grade Fund

14

Vanguard Intermediate-Term Treasury Index Fund

11

Vanguard Russell 1000 Growth Index Fund

10

Vanguard Real Estate Index Fund

14

Vanguard Precious Metals and Mining Fund

14

Vanguard Real Estate II Index Fund

9

Vanguard Long-Term Treasury Fund

25

Vanguard Long-Term Treasury Index Fund

10

Vanguard Russell 1000 Index Fund

10

Vanguard Russell 1000 Value Index Fund

10

Vanguard Russell 2000 Value Index Fund

10

Vanguard Russell 2000 Index Fund

10

Vanguard Russell 2000 Growth Index Fund

11

Vanguard Russell 3000 Index Fund

10

Vanguard Short-Term Treasury Index Fund

11

Vanguard Ultra-Short-Term Bond Fund

14

```
Vanguard Short-Term Federal Fund
11
Vanguard Short-Term Corporate Bond Index Fund
12
Vanguard Short-Term Investment-Grade Fund
13
Vanguard Short-Term Treasury Fund
10
Variable Portfolio - BlackRock Global Inflation-Protected Securities Fund
46
Variable Portfolio - MFS Blended Research Core Equity Fund
30
Variable Portfolio - MFS Value Fund
24
Variable Portfolio - Wells Fargo Short Duration Government Fund
39
Variable Portfolio - Morgan Stanley Advantage Fund
29
Variable Portfolio - Partners Small Cap Value Fund
27
Variable Portfolio - T. Rowe Price Large Cap Value Fund
21
Variable Portfolio - Oppenheimer International Growth Fund
31
Virtus Enhanced U.S. Equity ETF
29
Virtus Newfleet Dynamic Credit ETF
56
Western Asset Intermediate Maturity California Municipals Fund
52
Variable Portfolio - Victory Sycamore Established Value Fund
23
Variable Portfolio - TCW Core Plus Bond Fund
53
Western Asset Intermediate Maturity New York Municipals Fund
53
Western Asset Massachusetts Municipals Fund
52
World Growth Fund
17
```

[93]: `pd.set_option('display.max_colwidth', None)`

[94]: `df_RAG`

[94]:
```
                                                         fund_name  \
0                                             AB Arizona Portfolio
```

```
1                           AB Asia ex-Japan Equity Portfolio
2                             AB Credit Long/Short Portfolio
3                        AB Balanced Wealth Strategy Portfolio
4                                  AB Bond Inflation Strategy
..                                                           …
540  Variable Portfolio - Victory Sycamore Established Value Fund
541                     Variable Portfolio - TCW Core Plus Bond Fund
542  Western Asset Intermediate Maturity New York Municipals Fund
543                   Western Asset Massachusetts Municipals Fund
544                                              World Growth Fund


          Predicted investment strategy  \
0    Fixed Income Long Only (Low Risk)
1            Equity Long Only (Low Risk)
2         Long Short Funds (High Risk)
3              Balanced Fund (Low Risk)
4    Fixed Income Long Only (Low Risk)
..                                    …
540          Equity Long Only (Low Risk)
541  Fixed Income Long Only (Low Risk)
542  Fixed Income Long Only (Low Risk)
543  Fixed Income Long Only (Low Risk)
544       Long Short Funds (High Risk)


                 investment strategy evidence  \
0
AB Arizona Portfolio\n\n<head_breaker>\n\nAB Arizona Portfolio\n\n-------------
----------------------------------------------------------------\n\nINVESTMENT
OBJECTIVE: The investment objective of the Portfolio is to earn the highest
level of current income exempt from both federal income tax and State of Arizona
personal income tax that is available without assuming what the Adviser
considers to be undue risk.\n\nFEES AND EXPENSES OF THE PORTFOLIO:
1
AB Asia ex\n\nJapan Equity Portfolio\n\n<head_breaker>\n\nAB ASIA EX\n\nJAPAN
EQUITY PORTFOLIO\n\n-----------------------------------------------------------
--------------------\n\nAT A RECENT MEETING, THE FUND'S BOARD OF DIRECTORS
APPROVED THE PROPOSED LIQUIDATION AND TERMINATION OF THE FUND (THE
"LIQUIDATION"). IN CONNECTION WITH THE LIQUIDATION, THE FUND HAS SUSPENDED MOST
SALES OF ITS SHARES AND EXCHANGES INTO THE FUND. IN LIMITED CIRCUMSTANCES, SUCH
AS SALES TO CERTAIN RETIREMENT PLANS AND SALES MADE THROUGH RETAIL OMNIBUS
PLATFORMS, THE FUND WILL CONTINUE TO OFFER ITS SHARES, ALTHOUGH WITHOUT AN
INITIAL SALES CHARGE. IN ADDITION, THE FUND HAS SUSPENDED DISTRIBUTION AND/OR
SERVICE (RULE 12B-1) FEES AND WAIVED CONTINGENT DEFERRED SALES CHARGES UPON
REDEMPTIONS OF THE FUND'S SHARES.\n\nINVESTMENT OBJECTIVE The Fund's investment
objective is to seek long-term growth of capital.\n\nFEES AND EXPENSES OF THE
FUND
2               AB Credit Long/Short Portfolio\n\n<head_breaker>\n\nAB CREDIT
```

LONG/SHORT PORTFOLIO\n\n---
-----------------------\n\nINVESTMENT OBJECTIVE The Fund's investment objective
is to seek absolute return over a full market cycle.\n\nFEES AND EXPENSES OF THE
FUND\n\nThis table describes the fees and expenses that you may pay if you buy
and hold shares of the Fund. You may qualify for sales charge reductions if you
and members of your family invest, or agree to invest in the future, at least
$100,000 in AB Mutual Funds. More information about these and other discounts is
available from your financial intermediary and in Investing in the Funds--Sales
Charge Reduction Programs for Class A Shares on page 61 of this Prospectus, in
Appendix C--Financial Intermediary Waivers of this Prospectus and in Purchase of
Shares--Sales Charge Reduction Programs for Class A Shares on page 126 of the
Fund's Statement of Additional Information ("SAI").

3   AB Balanced
Wealth Strategy Portfolio <head_breaker>\n\nAB VPS BALANCED WEALTH STRATEGY
PORTFOLIO --
----------\n\nINVESTMENT OBJECTIVE The Portfolio's investment objective is to
maximize total return consistent with the Adviser's determination of reasonable
risk.\n\nFEES AND EXPENSES OF THE PORTFOLIO\n\nThis table describes the fees and
expenses that you may pay if you buy and hold shares of the Portfolio. Because
the information does not reflect deductions at the separate account level or
contract level for any charges that may be incurred under a contract,
Contractholders that invest in the Portfolio should refer to the variable
contract prospectus for a description of fees and expenses that apply to
Contractholders. Inclusion of these charges would increase the fees and expenses
provided below.\n\nSHAREHOLDER FEES (fees paid directly from your investment)
N/A

4
AB Bond Inflation Strategy\n\n<head_breaker>\n\nSUMMARY INFORMATION\n\n---------
--\n\nAB
BOND INFLATION STRATEGY\n\n--
--------------------------\n\nINVESTMENT OBJECTIVE The Fund's investment
objective is to maximize real return without assuming what the Adviser considers
to be undue risk.\n\nFEES AND EXPENSES OF THE FUND
..
…
540
Variable Portfolio - Victory Sycamore Established Value Fund
<head_breaker>\n\nSUMMARY OF CTIVPSM - VICTORY SYCAMORE ESTABLISHED VALUE FUND
Investment Objective CTIVPSM - Victory Sycamore Established Value Fund (the
Fund) seeks to provide shareholders with long-term growth of capital. Fees and
Expenses of the Fund This table describes the fees and expenses that you may pay
as an investor in the Fund. The table does not reflect any fees or expenses
imposed by your Contract or Qualified Plan, which are disclosed in your separate
Contract prospectus or Qualified Plan disclosure documents. If the additional
fees or expenses were reflected, the expenses set forth below would be higher.
541
Variable Portfolio - TCW Core Plus Bond Fund <head_breaker>\n\nSUMMARY OF

CTIVPSM - TCW CORE PLUS BOND FUND Investment Objective CTIVPSM - TCW Core Plus Bond Fund (the Fund) seeks to provide shareholders with total return through current income and capital appreciation. Fees and Expenses of the Fund This table describes the fees and expenses that you may pay as an investor in the Fund. The table does not reflect any fees or expenses imposed by your Contract or Qualified Plan, which are disclosed in your separate Contract prospectus or Qualified Plan disclosure documents. If the additional fees or expenses were reflected, the expenses set forth below would be higher.

542 The fund may also invest in securities of other open-end or closed-end investment companies, including exchange-traded funds ("ETFs"), that invest primarily in municipal securities.\n\nInstead of, and/or in addition to, investing directly in particular securities, the fund may use instruments such as derivatives, including options, futures contracts and inverse floating rate instruments issued in tender option bond transactions, and other synthetic instruments that are intended to provide economic exposure to the securities or the issuer or to be used as a hedging technique. The fund may use one or more types of these\n\nWestern Asset Intermediate Maturity New York Municipals Fund\t \t3 Table of Contents instruments without limit, except that these instruments are taken into account when determining compliance with the fund's 80% policy. For additional information regarding derivatives, see "More on the fund's investment strategies, investments and risks-Derivatives" in this Prospectus.

543
The fund may also engage in a variety of transactions using derivatives in order to change the investment characteristics of its portfolio (such as shortening or lengthening duration) and for other purposes. The fund may leverage its assets by investing proceeds received through tender option\n\nWestern Asset Massachusetts Municipals Fund\t \t3 Table of Contents bond transactions, which is considered a form of borrowing. See "More on the fund's investment strategies, investments and risks-Tender option bonds" in this Prospectus.\n\nThe fund is classified as "non-diversified," which means it may invest a larger percentage of its assets in a smaller number of issuers than a diversified fund.\n\nPrincipal risks

544
World Growth Fund\n\n<head_breaker>

     source meta data
0   {'source': 'MutualFundSummary/AB
Arizona Portfolio.txt'}
1   {'source': 'MutualFundSummary/AB Asia ex-Japan
Equity Portfolio.txt'}
2   {'source': 'MutualFundSummary/AB Credit
LongShort Portfolio.txt'}
3   {'source': 'MutualFundSummary/AB Balanced Wealth
Strategy Portfolio.txt'}
4   {'source': 'MutualFundSummary/AB Bond
Inflation Strategy.txt'}
..

```
…
540  {'source': 'MutualFundSummary/Variable Portfolio - Victory Sycamore
Established Value Fund.txt'}
541                {'source': 'MutualFundSummary/Variable Portfolio - TCW Core
Plus Bond Fund.txt'}
542  {'source': 'MutualFundSummary/Western Asset Intermediate Maturity New York
Municipals Fund.txt'}
543                {'source': 'MutualFundSummary/Western Asset Massachusetts
Municipals Fund.txt'}
544                                            {'source':
'MutualFundSummary/World Growth Fund.txt'}

[545 rows x 4 columns]
```

```
[95]: df_RAG.to_csv('df_RAG_all.csv', index=False)
```

0.2 Create new folder for failed process

```
[117]: fail = df_RAG[33:]
       fail
```

```
[117]:                             fund_name Predicted investment strategy  \
       33  Fidelity Series Opportunistic Insights Fund                      NaN
       34        Fidelity Series Short-Term Credit Fund                      NaN
       35              Fidelity Simplicity RMD 2005 Fund                      NaN
       36              Fidelity Simplicity RMD 2020 Fund                      NaN
       37    Fidelity ZERO Extended Market Index Fund                        NaN
       38          Fidelity Simplicity RMD Income Fund                        NaN
       39          Fidelity ZERO Large Cap Index Fund                         NaN
       40              Franklin FTSE Asia ex Japan ETF                        NaN
       41            FOCUSED INTERNATIONAL GROWTH FUND                        NaN
       42                    Franklin FTSE Brazil ETF                         NaN
       43                  Franklin FTSE Australia ETF                        NaN
       44                     Franklin FTSE China ETF                         NaN
       45            Franklin FTSE Europe Hedged ETF                          NaN
       46                   Franklin FTSE Germany ETF                         NaN
       47                    Franklin FTSE Europe ETF                         NaN
       48                    Franklin FTSE Canada ETF                         NaN
       49                    Franklin FTSE France ETF                         NaN
       50                 Franklin FTSE Hong Kong ETF                         NaN
       51                     Franklin FTSE Italy ETF                         NaN
       52                     Franklin FTSE India ETF                         NaN
       53               Franklin FTSE Japan Hedged ETF                        NaN
       54                     Franklin FTSE Japan ETF                         NaN

           investment strategy evidence source meta data
       33                   NaN                      NaN
```

34	NaN	NaN
35	NaN	NaN
36	NaN	NaN
37	NaN	NaN
38	NaN	NaN
39	NaN	NaN
40	NaN	NaN
41	NaN	NaN
42	NaN	NaN
43	NaN	NaN
44	NaN	NaN
45	NaN	NaN
46	NaN	NaN
47	NaN	NaN
48	NaN	NaN
49	NaN	NaN
50	NaN	NaN
51	NaN	NaN
52	NaN	NaN
53	NaN	NaN
54	NaN	NaN

```
[119]: fund_name = fail['fund_name'].tolist()
       fund_name
```

```
[119]: ['Fidelity Series Opportunistic Insights Fund',
        'Fidelity Series Short-Term Credit Fund',
        'Fidelity Simplicity RMD 2005 Fund',
        'Fidelity Simplicity RMD 2020 Fund',
        'Fidelity ZERO Extended Market Index Fund',
        'Fidelity Simplicity RMD Income Fund',
        'Fidelity ZERO Large Cap Index Fund',
        'Franklin FTSE Asia ex Japan ETF',
        'FOCUSED INTERNATIONAL GROWTH FUND',
        'Franklin FTSE Brazil ETF',
        'Franklin FTSE Australia ETF',
        'Franklin FTSE China ETF',
        'Franklin FTSE Europe Hedged ETF',
        'Franklin FTSE Germany ETF',
        'Franklin FTSE Europe ETF',
        'Franklin FTSE Canada ETF',
        'Franklin FTSE France ETF',
        'Franklin FTSE Hong Kong ETF',
        'Franklin FTSE Italy ETF',
        'Franklin FTSE India ETF',
        'Franklin FTSE Japan Hedged ETF',
        'Franklin FTSE Japan ETF']
```

```
[121]: #add .txt
       fund_names_with_txt = [name + ".txt" for name in fund_name]
       fund_names_with_txt
```

```
[121]: ['Fidelity Series Opportunistic Insights Fund.txt',
        'Fidelity Series Short-Term Credit Fund.txt',
        'Fidelity Simplicity RMD 2005 Fund.txt',
        'Fidelity Simplicity RMD 2020 Fund.txt',
        'Fidelity ZERO Extended Market Index Fund.txt',
        'Fidelity Simplicity RMD Income Fund.txt',
        'Fidelity ZERO Large Cap Index Fund.txt',
        'Franklin FTSE Asia ex Japan ETF.txt',
        'FOCUSED INTERNATIONAL GROWTH FUND.txt',
        'Franklin FTSE Brazil ETF.txt',
        'Franklin FTSE Australia ETF.txt',
        'Franklin FTSE China ETF.txt',
        'Franklin FTSE Europe Hedged ETF.txt',
        'Franklin FTSE Germany ETF.txt',
        'Franklin FTSE Europe ETF.txt',
        'Franklin FTSE Canada ETF.txt',
        'Franklin FTSE France ETF.txt',
        'Franklin FTSE Hong Kong ETF.txt',
        'Franklin FTSE Italy ETF.txt',
        'Franklin FTSE India ETF.txt',
        'Franklin FTSE Japan Hedged ETF.txt',
        'Franklin FTSE Japan ETF.txt']
```

```
[122]: len(fund_names_with_txt)
```

```
[122]: 22
```

```
[123]: %cd /content/gdrive/MyDrive/'Colab Notebooks'/MF815/NLP/'NLP_app '/
       ↪MutualFundSummary4
```

/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app /MutualFundSummary4

```
[124]: !mkdir -p MutualFundSummary4_5
```

```
[125]: #move failed file to new folder
       !mv \
       'Fidelity Series Opportunistic Insights Fund.txt' \
       'Fidelity Series Short-Term Credit Fund.txt' \
       'Fidelity Simplicity RMD 2005 Fund.txt' \
       'Fidelity Simplicity RMD 2020 Fund.txt' \
       'Fidelity ZERO Extended Market Index Fund.txt' \
       'Fidelity Simplicity RMD Income Fund.txt' \
       'Fidelity ZERO Large Cap Index Fund.txt' \
```

```
'Franklin FTSE Asia ex Japan ETF.txt' \
'FOCUSED INTERNATIONAL GROWTH FUND.txt' \
'Franklin FTSE Brazil ETF.txt' \
'Franklin FTSE Australia ETF.txt' \
'Franklin FTSE China ETF.txt' \
'Franklin FTSE Europe Hedged ETF.txt' \
'Franklin FTSE Germany ETF.txt' \
'Franklin FTSE Europe ETF.txt' \
'Franklin FTSE Canada ETF.txt' \
'Franklin FTSE France ETF.txt' \
'Franklin FTSE Hong Kong ETF.txt' \
'Franklin FTSE Italy ETF.txt' \
'Franklin FTSE India ETF.txt' \
'Franklin FTSE Japan Hedged ETF.txt' \
'Franklin FTSE Japan ETF.txt' \
MutualFundSummary4_5/
```

[126]:
```
%cd /content/gdrive/MyDrive/'Colab Notebooks'/MF815/NLP/'NLP_app '
```

/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app

[ ]:

RAG_Comparison

May 3, 2025

0.1 RAG Label Comparison

```python
import pandas as pd
import numpy as np
```

```python
def replace_with_value(value):
    if value.startswith('Balanced Fund'):
        return 2
    elif value.startswith('Fixed Income'):
        return 1
    elif value.startswith('Equity Long Only'):
        return 0
    else:
        return np.nan
    return value
```

```python
df_RAG = pd.read_csv("df_RAG_all.csv")
df_RAG
```

```
[6]:                                            fund_name  \
     0                            AB Arizona Portfolio
     1                  AB Asia ex-Japan Equity Portfolio
     2                       AB Credit Long/Short Portfolio
     3                AB Balanced Wealth Strategy Portfolio
     4                           AB Bond Inflation Strategy
     ..                                                 …
     540  Variable Portfolio - Victory Sycamore Establis…
     541      Variable Portfolio - TCW Core Plus Bond Fund
     542  Western Asset Intermediate Maturity New York M…
     543        Western Asset Massachusetts Municipals Fund
     544                                   World Growth Fund

               Predicted investment strategy  \
     0    Fixed Income Long Only (Low Risk)
     1          Equity Long Only (Low Risk)
     2         Long Short Funds (High Risk)
     3            Balanced Fund (Low Risk)
     4    Fixed Income Long Only (Low Risk)
```

```
..                                                     …
540            Equity Long Only (Low Risk)
541   Fixed Income Long Only (Low Risk)
542   Fixed Income Long Only (Low Risk)
543   Fixed Income Long Only (Low Risk)
544            Long Short Funds (High Risk)


                              investment strategy evidence  \
0      AB Arizona Portfolio\n\n<head_breaker>\n\nAB A…
1      AB Asia ex\n\nJapan Equity Portfolio\n\n<head_…
2      AB Credit Long/Short Portfolio\n\n<head_breake…
3      AB Balanced Wealth Strategy Portfolio <head_br…
4      AB Bond Inflation Strategy\n\n<head_breaker>\n…
..                                                   …
540    Variable Portfolio - Victory Sycamore Establis…
541    Variable Portfolio - TCW Core Plus Bond Fund <…
542    The fund may also invest in securities of othe…
543    The fund may also engage in a variety of trans…
544               World Growth Fund\n\n<head_breaker>


                                 source meta data
0      {'source': 'MutualFundSummary/AB Arizona Portf…
1      {'source': 'MutualFundSummary/AB Asia ex-Japan…
2      {'source': 'MutualFundSummary/AB Credit LongSh…
3      {'source': 'MutualFundSummary/AB Balanced Weal…
4      {'source': 'MutualFundSummary/AB Bond Inflatio…
..                                                 …
540    {'source': 'MutualFundSummary/Variable Portfol…
541    {'source': 'MutualFundSummary/Variable Portfol…
542    {'source': 'MutualFundSummary/Western Asset In…
543    {'source': 'MutualFundSummary/Western Asset Ma…
544    {'source': 'MutualFundSummary/World Growth Fun…

[545 rows x 4 columns]
```

```python
df = pd.read_csv("MutualFundLabels.csv")
df
```

```
[7]:                      id                              fund_name  \
0    0000051931-18-000151           American Funds College 2018 Fund
1    0000051931-18-000151           American Funds College 2021 Fund
2    0000051931-18-000151           American Funds College 2024 Fund
3    0000051931-18-000151           American Funds College 2027 Fund
4    0000051931-18-000151           American Funds College 2030 Fund
..                    …                                          …
462  0001710607-18-000172   American Century Diversified Corporate Bond ETF
463  0001710607-18-000172   American Century Diversified Municipal Bond ETF
```

```
464   0001710607-18-000172     American Century Quality Diversified Internati…
465   0001710607-18-000172       American Century STOXX U.S. Quality Growth ETF
466   0001710607-18-000172        American Century STOXX U.S. Quality Value ETF


       Performance fee?                      Ivestment Strategy Leverage?  \
0                    NaN              Balanced Fund (Low Risk)       Yes
1                    NaN              Balanced Fund (Low Risk)       Yes
2                    NaN              Balanced Fund (Low Risk)       Yes
3                    NaN              Balanced Fund (Low Risk)       Yes
4                    NaN              Balanced Fund (Low Risk)       Yes
..                   …                                    …         …
462                  NaN    Fixed Income Long Only (Low Risk)       Yes
463                  NaN    Fixed Income Long Only (Low Risk)        No
464                  NaN          Equity Long Only (Low Risk)       Yes
465                  NaN          Equity Long Only (Low Risk)       Yes
466                  NaN          Equity Long Only (Low Risk)       Yes


                                  Portfolio composition   Concentration
0                          Investment grade securities    Diversified
1                          Investment grade securities    Diversified
2                          Investment grade securities    Diversified
3                          Investment grade securities    Diversified
4                          Investment grade securities    Diversified
..                                                   …              …
462                        Investment grade securities    Diversified
463                        Investment grade securities    Diversified
464   Sub-investment grade securities or emerging ma…    Diversified
465                                    Listed Equities    Diversified
466                                    Listed Equities    Diversified

[467 rows x 7 columns]
```

```
[8]:  df_RAG['Prediction'] = df_RAG['Predicted investment strategy'].
        ↪apply(replace_with_value)
      df['Ivestment_Strategy_label'] = df['Ivestment Strategy'].
        ↪apply(replace_with_value)
```

```
[9]:  df_RAG
```

```
[9]:                                              fund_name  \
      0                             AB Arizona Portfolio
      1                   AB Asia ex-Japan Equity Portfolio
      2                       AB Credit Long/Short Portfolio
      3                 AB Balanced Wealth Strategy Portfolio
      4                           AB Bond Inflation Strategy
      ..                                                 …
      540   Variable Portfolio – Victory Sycamore Establis…
```

```
541                  Variable Portfolio - TCW Core Plus Bond Fund
542        Western Asset Intermediate Maturity New York M…
543               Western Asset Massachusetts Municipals Fund
544                                           World Growth Fund

            Predicted investment strategy  \
0       Fixed Income Long Only (Low Risk)
1           Equity Long Only (Low Risk)
2          Long Short Funds (High Risk)
3               Balanced Fund (Low Risk)
4       Fixed Income Long Only (Low Risk)
..                                    …
540         Equity Long Only (Low Risk)
541     Fixed Income Long Only (Low Risk)
542     Fixed Income Long Only (Low Risk)
543     Fixed Income Long Only (Low Risk)
544         Long Short Funds (High Risk)

                           investment strategy evidence  \
0       AB Arizona Portfolio\n\n<head_breaker>\n\nAB A…
1       AB Asia ex\n\nJapan Equity Portfolio\n\n<head_…
2       AB Credit Long/Short Portfolio\n\n<head_breake…
3       AB Balanced Wealth Strategy Portfolio <head_br…
4       AB Bond Inflation Strategy\n\n<head_breaker>\n…
..                                                    …
540     Variable Portfolio - Victory Sycamore Establis…
541     Variable Portfolio - TCW Core Plus Bond Fund <…
542     The fund may also invest in securities of othe…
543     The fund may also engage in a variety of trans…
544                  World Growth Fund\n\n<head_breaker>

                          source meta data  Prediction
0       {'source': 'MutualFundSummary/AB Arizona Portf…         1.0
1       {'source': 'MutualFundSummary/AB Asia ex-Japan…         0.0
2       {'source': 'MutualFundSummary/AB Credit LongSh…         NaN
3       {'source': 'MutualFundSummary/AB Balanced Weal…         2.0
4       {'source': 'MutualFundSummary/AB Bond Inflatio…         1.0
..                                                    …           …
540     {'source': 'MutualFundSummary/Variable Portfol…         0.0
541     {'source': 'MutualFundSummary/Variable Portfol…         1.0
542     {'source': 'MutualFundSummary/Western Asset In…         1.0
543     {'source': 'MutualFundSummary/Western Asset Ma…         1.0
544     {'source': 'MutualFundSummary/World Growth Fun…         NaN

[545 rows x 5 columns]
```

[10]: `df`

```
[10]:                        id                                    fund_name  \
     0    0000051931-18-000151              American Funds College 2018 Fund
     1    0000051931-18-000151              American Funds College 2021 Fund
     2    0000051931-18-000151              American Funds College 2024 Fund
     3    0000051931-18-000151              American Funds College 2027 Fund
     4    0000051931-18-000151              American Funds College 2030 Fund
     ..                     …                                            …
     462  0001710607-18-000172     American Century Diversified Corporate Bond ETF
     463  0001710607-18-000172     American Century Diversified Municipal Bond ETF
     464  0001710607-18-000172   American Century Quality Diversified Internati…
     465  0001710607-18-000172       American Century STOXX U.S. Quality Growth ETF
     466  0001710607-18-000172        American Century STOXX U.S. Quality Value ETF

          Performance fee?              Ivestment Strategy Leverage?  \
     0               NaN            Balanced Fund (Low Risk)       Yes
     1               NaN            Balanced Fund (Low Risk)       Yes
     2               NaN            Balanced Fund (Low Risk)       Yes
     3               NaN            Balanced Fund (Low Risk)       Yes
     4               NaN            Balanced Fund (Low Risk)       Yes
     ..                …                            …           …
     462             NaN  Fixed Income Long Only (Low Risk)        Yes
     463             NaN  Fixed Income Long Only (Low Risk)         No
     464             NaN         Equity Long Only (Low Risk)       Yes
     465             NaN         Equity Long Only (Low Risk)       Yes
     466             NaN         Equity Long Only (Low Risk)       Yes

                                    Portfolio composition  Concentration  \
     0                        Investment grade securities     Diversified
     1                        Investment grade securities     Diversified
     2                        Investment grade securities     Diversified
     3                        Investment grade securities     Diversified
     4                        Investment grade securities     Diversified
     ..                                               …              …
     462                      Investment grade securities     Diversified
     463                      Investment grade securities     Diversified
     464  Sub-investment grade securities or emerging ma…     Diversified
     465                                 Listed Equities     Diversified
     466                                 Listed Equities     Diversified

          Ivestment_Strategy_label
     0                         2.0
     1                         2.0
     2                         2.0
     3                         2.0
     4                         2.0
     ..                         …
     462                       1.0
```

```
463                      1.0
464                      0.0
465                      0.0
466                      0.0

[467 rows x 8 columns]
```

```
[11]:  df_merged = df.merge(df_RAG, on='fund_name', how='inner')
       df_merged
```

```
[11]:                    id                                   fund_name  \
       0     0000051931-18-000151           American Funds College 2018 Fund
       1     0000051931-18-000151           American Funds College 2021 Fund
       2     0000051931-18-000151           American Funds College 2024 Fund
       3     0000051931-18-000151           American Funds College 2027 Fund
       4     0000051931-18-000151           American Funds College 2030 Fund
       ..                   ...                                         ...
       461   0001710607-18-000172    American Century Diversified Corporate Bond ETF
       462   0001710607-18-000172    American Century Diversified Municipal Bond ETF
       463   0001710607-18-000172  American Century Quality Diversified Internati…
       464   0001710607-18-000172     American Century STOXX U.S. Quality Growth ETF
       465   0001710607-18-000172      American Century STOXX U.S. Quality Value ETF

            Performance fee?              Ivestment Strategy Leverage?  \
       0                 NaN         Balanced Fund (Low Risk)        Yes
       1                 NaN         Balanced Fund (Low Risk)        Yes
       2                 NaN         Balanced Fund (Low Risk)        Yes
       3                 NaN         Balanced Fund (Low Risk)        Yes
       4                 NaN         Balanced Fund (Low Risk)        Yes
       ..                ...                              ...        ...
       461               NaN  Fixed Income Long Only (Low Risk)       Yes
       462               NaN  Fixed Income Long Only (Low Risk)        No
       463               NaN         Equity Long Only (Low Risk)      Yes
       464               NaN         Equity Long Only (Low Risk)      Yes
       465               NaN         Equity Long Only (Low Risk)      Yes

                                    Portfolio composition  Concentration  \
       0                        Investment grade securities    Diversified
       1                        Investment grade securities    Diversified
       2                        Investment grade securities    Diversified
       3                        Investment grade securities    Diversified
       4                        Investment grade securities    Diversified
       ..                                              ...            ...
       461                      Investment grade securities    Diversified
       462                      Investment grade securities    Diversified
       463  Sub-investment grade securities or emerging ma…    Diversified
       464                                  Listed Equities    Diversified
```

```
465                                 Listed Equities    Diversified

        Ivestment_Strategy_label      Predicted investment strategy  \
0                            2.0              Balanced Fund (Low Risk)
1                            2.0              Balanced Fund (Low Risk)
2                            2.0              Balanced Fund (Low Risk)
3                            2.0              Balanced Fund (Low Risk)
4                            2.0              Balanced Fund (Low Risk)
..                           ...                                   ...
461                          1.0  Fixed Income Long Only (Low Risk)
462                          1.0  Fixed Income Long Only (Low Risk)
463                          0.0         Equity Long Only (Low Risk)
464                          0.0         Equity Long Only (Low Risk)
465                          0.0         Equity Long Only (Low Risk)

                               investment strategy evidence  \
0       American Funds College 2018 Fund <head_breaker…
1       American Funds College 2021 Fund <head_breaker…
2       American Funds College 2024 Fund <head_breaker…
3       American Funds College 2027 Fund <head_breaker…
4       American Funds College 2030 Fund <head_breaker…
..                                                   …
461     American Century Diversified Corporate Bond ET…
462     American Century Diversified Municipal Bond ET…
463     American Century Quality Diversified Internati…
464     American Century STOXX U.S. Quality Growth ETF…
465     American Century STOXX U.S. Quality Value ETF …

                                 source meta data  Prediction
0       {'source': 'MutualFundSummary/American Funds C…         2.0
1       {'source': 'MutualFundSummary/American Funds C…         2.0
2       {'source': 'MutualFundSummary1/American Funds …         2.0
3       {'source': 'MutualFundSummary1/American Funds …         2.0
4       {'source': 'MutualFundSummary/American Funds C…         2.0
..                                                   …          …
461     {'source': 'MutualFundSummary/American Century…         1.0
462     {'source': 'MutualFundSummary/American Century…         1.0
463     {'source': 'MutualFundSummary/American Century…         0.0
464     {'source': 'MutualFundSummary/American Century…         0.0
465     {'source': 'MutualFundSummary/American Century…         0.0

[466 rows x 12 columns]
```

[12]:
```python
from sklearn.metrics import confusion_matrix
```

[13]:
```python
df_summary = df_merged.dropna(subset =
    ['Ivestment_Strategy_label','Prediction'])
```

```
[14]: df_summary

[14]:                        id                                 fund_name  \
      0     0000051931-18-000151           American Funds College 2018 Fund
      1     0000051931-18-000151           American Funds College 2021 Fund
      2     0000051931-18-000151           American Funds College 2024 Fund
      3     0000051931-18-000151           American Funds College 2027 Fund
      4     0000051931-18-000151           American Funds College 2030 Fund
      ..                      …                                         …
      461   0001710607-18-000172      American Century Diversified Corporate Bond ETF
      462   0001710607-18-000172      American Century Diversified Municipal Bond ETF
      463   0001710607-18-000172   American Century Quality Diversified Internati…
      464   0001710607-18-000172       American Century STOXX U.S. Quality Growth ETF
      465   0001710607-18-000172        American Century STOXX U.S. Quality Value ETF


            Performance fee?              Ivestment Strategy Leverage?  \
      0               NaN            Balanced Fund (Low Risk)       Yes
      1               NaN            Balanced Fund (Low Risk)       Yes
      2               NaN            Balanced Fund (Low Risk)       Yes
      3               NaN            Balanced Fund (Low Risk)       Yes
      4               NaN            Balanced Fund (Low Risk)       Yes
      ..               …                          …             …
      461             NaN  Fixed Income Long Only (Low Risk)       Yes
      462             NaN  Fixed Income Long Only (Low Risk)        No
      463             NaN       Equity Long Only (Low Risk)       Yes
      464             NaN       Equity Long Only (Low Risk)       Yes
      465             NaN       Equity Long Only (Low Risk)       Yes


                              Portfolio composition  Concentration  \
      0                 Investment grade securities    Diversified
      1                 Investment grade securities    Diversified
      2                 Investment grade securities    Diversified
      3                 Investment grade securities    Diversified
      4                 Investment grade securities    Diversified
      ..                                        …             …
      461               Investment grade securities    Diversified
      462               Investment grade securities    Diversified
      463   Sub-investment grade securities or emerging ma…    Diversified
      464                          Listed Equities    Diversified
      465                          Listed Equities    Diversified


            Ivestment_Strategy_label     Predicted investment strategy  \
      0                       2.0             Balanced Fund (Low Risk)
      1                       2.0             Balanced Fund (Low Risk)
      2                       2.0             Balanced Fund (Low Risk)
      3                       2.0             Balanced Fund (Low Risk)
      4                       2.0             Balanced Fund (Low Risk)
```

8

```
     ..                          ...                                    ...
     461                         1.0    Fixed Income Long Only (Low Risk)
     462                         1.0    Fixed Income Long Only (Low Risk)
     463                         0.0          Equity Long Only (Low Risk)
     464                         0.0          Equity Long Only (Low Risk)
     465                         0.0          Equity Long Only (Low Risk)

                                        investment strategy evidence  \
     0       American Funds College 2018 Fund <head_breaker…
     1       American Funds College 2021 Fund <head_breaker…
     2       American Funds College 2024 Fund <head_breaker…
     3       American Funds College 2027 Fund <head_breaker…
     4       American Funds College 2030 Fund <head_breaker…
     ..                                                   …
     461     American Century Diversified Corporate Bond ET…
     462     American Century Diversified Municipal Bond ET…
     463     American Century Quality Diversified Internati…
     464     American Century STOXX U.S. Quality Growth ETF…
     465     American Century STOXX U.S. Quality Value ETF …

                                        source meta data  Prediction
     0       {'source': 'MutualFundSummary/American Funds C…         2.0
     1       {'source': 'MutualFundSummary/American Funds C…         2.0
     2       {'source': 'MutualFundSummary1/American Funds …         2.0
     3       {'source': 'MutualFundSummary1/American Funds …         2.0
     4       {'source': 'MutualFundSummary/American Funds C…         2.0
     ..                                                   …           …
     461     {'source': 'MutualFundSummary/American Century…         1.0
     462     {'source': 'MutualFundSummary/American Century…         1.0
     463     {'source': 'MutualFundSummary/American Century…         0.0
     464     {'source': 'MutualFundSummary/American Century…         0.0
     465     {'source': 'MutualFundSummary/American Century…         0.0

     [432 rows x 12 columns]
```

```python
[15]: cm = confusion_matrix(df_summary['Ivestment_Strategy_label'],
       df_summary['Prediction'])
      print(cm)
```

```
[[205   1  16]
 [  8 118   3]
 [  4   2  75]]
```

```python
[16]: import seaborn as sns
      import matplotlib.pyplot as plt

      sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
```

```
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```
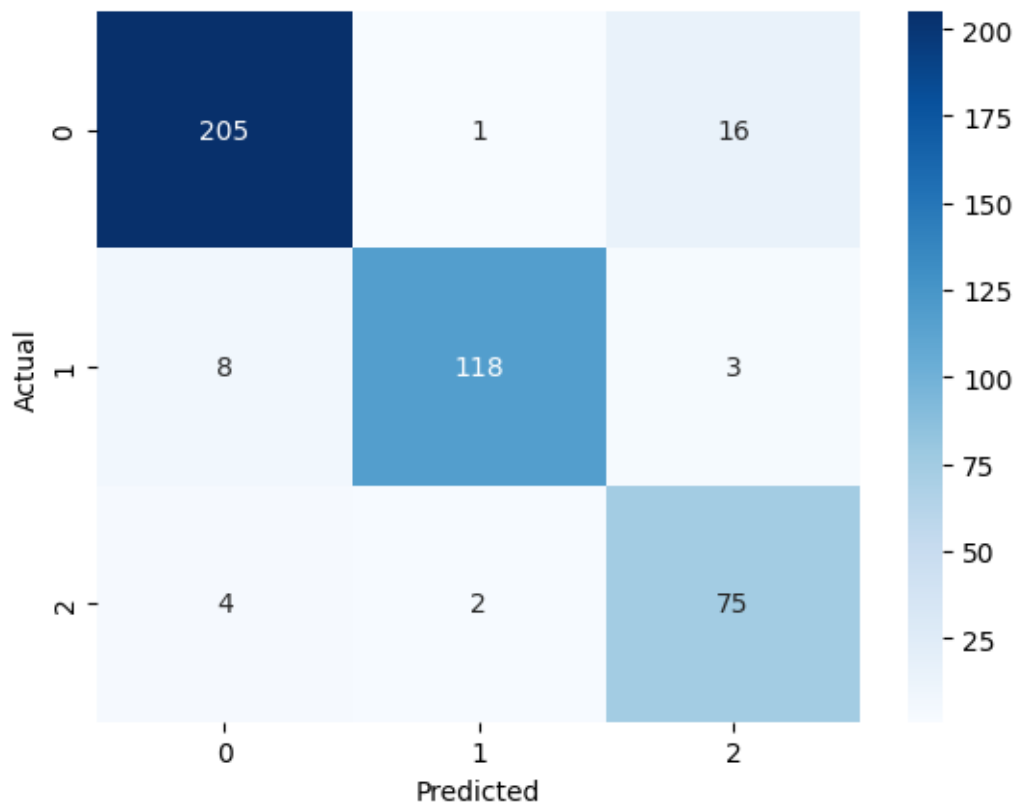


```
[17]: df_comparison = df_summary[['fund_name',␣
       ↪'Ivestment_Strategy_label','Prediction','investment strategy␣
       ↪evidence','source meta data']]
```

```
[18]: df_comparison = df_comparison[df_comparison['Ivestment_Strategy_label'] !=␣
       ↪df_comparison['Prediction']]
```

```
[19]: df_comparison.shape
```

```
[19]: (34, 5)
```

```
[20]: #Pick several funds where the labels from these two sources are not the same
      df_comparison_sample = df_comparison.sample(3,random_state=815)
      df_comparison_sample
```

```
[20]:                              fund_name  Ivestment_Strategy_label  \
      17           Global Growth and Income Fund                       0.0
```

```
221   Janus Henderson Global Research Portfolio                    1.0
23          International Growth and Income Fund                    0.0

       Prediction                      investment strategy evidence  \
17           2.0  Global Growth and Income Fund <head_breaker>\n…
221          0.0  Janus Henderson Global Research Portfolio <hea…
23           2.0  International Growth and Income Fund <head_bre…

                                       source meta data
17   {'source': 'MutualFundSummary/Global Growth an…
221  {'source': 'MutualFundSummary/Janus Henderson …
23   {'source': 'MutualFundSummary/International Gr…
```

[21]:
```python
df_comparison_sample['source meta data'].iloc[0], df_comparison_sample['source␣
 ↪meta data'].iloc[1],\
df_comparison_sample['source meta data'].iloc[2]
```

[21]:
```
("{'source': 'MutualFundSummary/Global Growth and Income Fund.txt'}",
 "{'source': 'MutualFundSummary/Janus Henderson Global Research
 Portfolio.txt'}",
 "{'source': 'MutualFundSummary/International Growth and Income Fund.txt'}")
```

[22]:
```python
df_comparison_sample['investment strategy evidence'].iloc[0]
```

[22]:
```
'Global Growth and Income Fund <head_breaker>\n\nGlobal Growth and Income
Fund\n\nInvestment objective The fund's investment objective is to provide long-
term growth of capital while providing current income.\n\nFees and expenses of
the fund This table describes the fees and expenses that you may pay if you buy
and hold an interest in Class 1 shares of the fund. It does not reflect
insurance contract fees and expenses. If insurance contract fees and expenses
were reflected, expenses shown would be higher.\n\nAnnual fund operating
expenses (expenses that you pay each year as a percentage of the value of your
investment)\tClass 1 Management fee\t0.59% Other expenses\t0.04 Total annual
fund operating expenses\t0.63 Example This example is intended to help you
compare the cost of investing in Class 1 shares of the fund with the cost of
investing in other mutual funds.'
```

[23]:
```python
df_comparison_sample['investment strategy evidence'].iloc[1]
```

[23]:
```
'Janus Henderson Global Research Portfolio <head_breaker>\n\nJanus Henderson
Global Research Portfolio Ticker:\tJAWGX\tInstitutional Shares'
```

[24]:
```python
df_comparison_sample['investment strategy evidence'].iloc[2]
```

[24]:
```
'International Growth and Income Fund <head_breaker>\n\nInternational Growth and
Income Fund\n\nInvestment objective The fund's investment objective is to
provide long-term growth of capital while providing current income.\n\nFees and
```

expenses of the fund This table describes the fees and expenses that you may pay
if you buy and hold an interest in Class 1 shares of the fund. It does not
reflect insurance contract fees and expenses. If insurance contract fees and
expenses were reflected, expenses shown would be higher.\n\nAnnual fund
operating expenses (expenses that you pay each year as a percentage of the value
of your investment)\tClass 1 Management fee\t0.61% Other expenses\t0.04 Total
annual fund operating expenses\t0.65 Example This example is intended to help
you compare the cost of investing in Class 1 shares of the fund with the cost of
investing in other mutual funds.'

Comment: First of all, our RAG process was able to retrieve appropriate txt files. Looking at the
first fund (Global Growth and Income Fund), the evidence provided by the RAG process shows that
"the fund's investment objective is to provide long-term growth of capital while providing current
income". It also mentioned interest. From those evidence, we can say this fund is using equity and
fixed income, we believe the answer provided by RAG is more accurate. For second fund (Janus
Henderson Global Research Portfolio), it looks the evidence provided from RAG is nothing t do
with investment strategy, so we think the labled result is more accurate. Lastly, for International
Growth and Income Fund, the fund's objective is similar to Global Growth and Income Fund. From
the evidence we have provided, we conclude this fund trade multiple asset, including fixed income
and equity, so the pridiction is more accurate

[25]: df_comparison

[25]:
```
                                            fund_name  \
17                      Global Growth and Income Fund
20                                  Growth-Income Fund
23               International Growth and Income Fund
25      Managed Risk Blue Chip Income and Growth Fund
26                            Managed Risk Growth Fund
27                     Managed Risk Growth-Income Fund
28                     Managed Risk International Fund
30                                      New World Fund
32   Portfolio Series - American Funds Growth and I…
34   Portfolio Series - American Funds Managed Risk…
35   Portfolio Series - American Funds Managed Risk…
96              Vanguard Russell 1000 Growth Index Fund
97                     Vanguard Russell 1000 Index Fund
98               Vanguard Russell 1000 Value Index Fund
99              Vanguard Russell 2000 Growth Index Fund
100                    Vanguard Russell 2000 Index Fund
101             Vanguard Russell 2000 Value Index Fund
102                    Vanguard Russell 3000 Index Fund
126                        MFS Global Total Return Fund
155             MainStay VP MFS Utilities Portfolio
174             Janus Henderson Growth and Income Fund
221             Janus Henderson Global Research Portfolio
223  Janus Henderson Global Unconstrained Bond Port…
```

```
225                  Janus Henderson Overseas Portfolio
254                      AllianzGI Income & Growth Fund
273  Ivy LaSalle Global Risk-Managed Real Estate Fund
276       Ivy Managed International Opportunities Fund
285           Ivy Securian Real Estate Securities Fund
291                          NORTHERN MONEY MARKET FUND
390         Fidelity Managed Retirement Income Fund
395            Fidelity Simplicity RMD Income Fund
398                          Franklin Payout 2020 Fund
399                          Franklin Payout 2021 Fund
419                              Swan Defined Risk Fund

     Ivestment_Strategy_label  Prediction  \
17                        0.0         2.0
20                        0.0         2.0
23                        0.0         2.0
25                        0.0         2.0
26                        0.0         2.0
27                        0.0         2.0
28                        0.0         2.0
30                        2.0         0.0
32                        0.0         2.0
34                        0.0         2.0
35                        0.0         2.0
96                        1.0         0.0
97                        1.0         0.0
98                        1.0         0.0
99                        1.0         0.0
100                       1.0         0.0
101                       1.0         0.0
102                       1.0         0.0
126                       0.0         2.0
155                       2.0         0.0
174                       0.0         2.0
221                       1.0         0.0
223                       0.0         1.0
225                       2.0         0.0
254                       0.0         2.0
273                       0.0         2.0
276                       2.0         0.0
285                       0.0         2.0
291                       1.0         2.0
390                       2.0         1.0
395                       2.0         1.0
398                       1.0         2.0
399                       1.0         2.0
419                       0.0         2.0
```

```
                         investment strategy evidence  \
17   Global Growth and Income Fund <head_breaker>\n…
20   Growth\n\nIncome Fund\n\n<head_breaker>\n\nGro…
23   International Growth and Income Fund <head_bre…
25   Managed Risk Blue Chip Income and Growth Fund …
26   Managed Risk Growth Fund\n\n<head_breaker>\n\n…
27   Managed Risk Growth\n\nIncome Fund\n\n<head_br…
28   Managed Risk International Fund\n\n<head_break…
30   New World Fund\n\n<head_breaker>\n\nNew World …
32   Portfolio Series - American Funds Growth and I…
34   Portfolio Series - American Funds Managed Risk…
35   Portfolio Series - American Funds Managed Risk…
96   Vanguard Russell 1000 Growth Index Fund <head_…
97   Vanguard Russell 1000 Index Fund <head_breaker…
98   Vanguard Russell 1000 Value Index Fund <head_b…
99   Vanguard Russell 2000 Growth Index Fund <head_…
100  Vanguard Russell 2000 Index Fund <head_breaker…
101  Vanguard Russell 2000 Value Index Fund <head_b…
102  Vanguard Russell 3000 Index Fund <head_breaker…
126  MFS Global Total Return Fund <head_breaker>\n\…
155  50\n\nMainStay VP MFS® Utilities Portfolio\n\n…
174  Janus Henderson Growth and Income Fund <head_b…
221  Janus Henderson Global Research Portfolio <hea…
223  Janus Henderson Global Unconstrained Bond Port…
225  Janus Henderson Overseas Portfolio\n\n<head_br…
254  AllianzGI Income & Growth Fund <head_breaker>\…
273  Principal Investment Strategies\n\nIvy LaSalle…
276  Principal Investment Strategies\n\nIvy Managed…
285  Principal Investment Strategies\n\nIvy Securia…
291  NORTHERN MONEY MARKET FUND\n\n<head_breaker>\n…
390  Fidelity Managed Retirement Income Fund <head_…
395  Fidelity Simplicity RMD Income Fund <head_brea…
398  Franklin Payout 2020 Fund\n\n<head_breaker>\n\…
399  Franklin Payout 2021 Fund\n\n<head_breaker>\n\…
419  Swan Defined Risk Fund\n\n<head_breaker>\n\nFU…

                                       source meta data
17   {'source': 'MutualFundSummary/Global Growth an…
20   {'source': 'MutualFundSummary/Growth-Income Fu…
23   {'source': 'MutualFundSummary/International Gr…
25   {'source': 'MutualFundSummary/Managed Risk Blu…
26   {'source': 'MutualFundSummary/Managed Risk Gro…
27   {'source': 'MutualFundSummary/Managed Risk Gro…
28   {'source': 'MutualFundSummary/Managed Risk Int…
30   {'source': 'MutualFundSummary/New World Fund.t…
32   {'source': 'MutualFundSummary/Portfolio Series…
```

```
34   {'source': 'MutualFundSummary/Portfolio Series…
35   {'source': 'MutualFundSummary/Portfolio Series…
96   {'source': 'MutualFundSummary/Vanguard Russell…
97   {'source': 'MutualFundSummary/Vanguard Russell…
98   {'source': 'MutualFundSummary/Vanguard Russell…
99   {'source': 'MutualFundSummary/Vanguard Russell…
100  {'source': 'MutualFundSummary/Vanguard Russell…
101  {'source': 'MutualFundSummary/Vanguard Russell…
102  {'source': 'MutualFundSummary/Vanguard Russell…
126  {'source': 'MutualFundSummary/MFS Global Total…
155  {'source': 'MutualFundSummary/MainStay VP MFS …
174  {'source': 'MutualFundSummary/Janus Henderson …
221  {'source': 'MutualFundSummary/Janus Henderson …
223  {'source': 'MutualFundSummary/Janus Henderson …
225  {'source': 'MutualFundSummary/Janus Henderson …
254  {'source': 'MutualFundSummary/AllianzGI Income…
273  {'source': 'MutualFundSummary/Ivy LaSalle Glob…
276  {'source': 'MutualFundSummary/Ivy Managed Inte…
285  {'source': 'MutualFundSummary/Ivy Securian Rea…
291  {'source': 'MutualFundSummary/NORTHERN MONEY M…
390  {'source': 'MutualFundSummary4/Fidelity Manage…
395  {'source': 'MutualFundSummary/Fidelity Simplic…
398  {'source': 'MutualFundSummary/Franklin Payout …
399  {'source': 'MutualFundSummary/Franklin Payout …
419  {'source': 'MutualFundSummary/Swan Defined Ris…
```

[26]: `df_summary.to_csv("df_summary.csv", index = False)`

0.2 Train Classification Algorithm

[ ]:

MF815_Final_Report_ver2

May 3, 2025

0.1 Install

```
[ ]: !pip install langchain
     !pip install langchain-chroma
     !pip install -U langchain-community
     !pip install tiktoken
     !pip install unstructured
```

Requirement already satisfied: langchain in /usr/local/lib/python3.11/dist-
packages (0.3.25)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.58 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.58)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.8 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.8)
Requirement already satisfied: langsmith<0.4,>=0.1.17 in
/usr/local/lib/python3.11/dist-packages (from langchain) (0.3.41)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in
/usr/local/lib/python3.11/dist-packages (from langchain) (2.11.3)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.11/dist-packages (from langchain) (2.0.40)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-
packages (from langchain) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-
packages (from langchain) (6.0.2)
Requirement already satisfied: tenacity!=8.4.0,<10.0.0,>=8.1.0 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.58->langchain) (9.1.2)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.58->langchain) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.58->langchain) (24.2)
Requirement already satisfied: typing-extensions>=4.7 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.58->langchain) (4.13.2)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)

(0.28.1)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(3.10.18)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from langsmith<0.4,>=0.1.17->langchain)
(0.23.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain)
(0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain)
(2.33.1)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic<3.0.0,>=2.7.4->langchain)
(0.4.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests<3,>=2->langchain) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain)
(2025.4.26)
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-
packages (from SQLAlchemy<3,>=1.4->langchain) (3.2.1)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain) (4.9.0)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain) (1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from
httpcore==1.*->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain) (0.16.0)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core<1.0.0,>=0.3.58->langchain) (3.0.0)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.17->langchain)
(1.3.1)
Requirement already satisfied: langchain-chroma in
/usr/local/lib/python3.11/dist-packages (0.2.3)
Requirement already satisfied: langchain-core>=0.3.52 in
/usr/local/lib/python3.11/dist-packages (from langchain-chroma) (0.3.58)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.11/dist-

packages (from langchain-chroma) (2.0.2)
Requirement already satisfied: chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from langchain-chroma) (0.6.3)
Requirement already satisfied: build>=1.0.3 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.2.2.post1)
Requirement already satisfied: pydantic>=1.9 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (2.11.3)
Requirement already satisfied: chroma-hnswlib==0.7.6 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.7.6)
Requirement already satisfied: fastapi>=0.95.2 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.115.12)
Requirement already satisfied: uvicorn>=0.18.3 in /usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.34.2)
Requirement already satisfied: posthog>=2.4.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (4.0.1)
Requirement already satisfied: typing_extensions>=4.5.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (4.13.2)
Requirement already satisfied: onnxruntime>=1.14.1 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.21.1)
Requirement already satisfied: opentelemetry-api>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.32.1)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-grpc>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.32.1)
Requirement already satisfied: opentelemetry-instrumentation-fastapi>=0.41b0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.53b1)
Requirement already satisfied: opentelemetry-sdk>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.32.1)
Requirement already satisfied: tokenizers>=0.13.2 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.21.1)
Requirement already satisfied: pypika>=0.48.9 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.48.9)
Requirement already satisfied: tqdm>=4.65.0 in /usr/local/lib/python3.11/dist-

packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (4.67.1)
Requirement already satisfied: overrides>=7.3.1 in
/usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (7.7.0)
Requirement already satisfied: importlib-resources in
/usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (6.5.2)
Requirement already satisfied: grpcio>=1.58.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.71.0)
Requirement already satisfied: bcrypt>=4.0.1 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (4.3.0)
Requirement already satisfied: typer>=0.9.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.15.3)
Requirement already satisfied: kubernetes>=28.1.0 in
/usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (32.0.1)
Requirement already satisfied: tenacity>=8.2.3 in
/usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (9.1.2)
Requirement already satisfied: PyYAML>=6.0.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (6.0.2)
Requirement already satisfied: mmh3>=4.0.1 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (5.1.0)
Requirement already satisfied: orjson>=3.9.12 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (3.10.18)
Requirement already satisfied: httpx>=0.27.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.28.1)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.11/dist-packages (from chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (14.0.0)
Requirement already satisfied: langsmith<0.4,>=0.1.125 in
/usr/local/lib/python3.11/dist-packages (from langchain-core>=0.3.52->langchain-chroma) (0.3.41)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-core>=0.3.52->langchain-chroma) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-core>=0.3.52->langchain-chroma) (24.2)
Requirement already satisfied: pyproject_hooks in

4

/usr/local/lib/python3.11/dist-packages (from build>=1.0.3->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.2.0)
Requirement already satisfied: starlette<0.47.0,>=0.40.0 in /usr/local/lib/python3.11/dist-packages (from fastapi>=0.95.2->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.46.2)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages (from httpx>=0.27.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (4.9.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from httpx>=0.27.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (2025.4.26)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-packages (from httpx>=0.27.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.0.9)
Requirement already satisfied: idna in /usr/local/lib/python3.11/dist-packages (from httpx>=0.27.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (3.10)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-packages (from httpcore==1.*->httpx>=0.27.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (0.16.0)
Requirement already satisfied: jsonpointer>=1.9 in /usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-core>=0.3.52->langchain-chroma) (3.0.0)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.17.0)
Requirement already satisfied: python-dateutil>=2.5.3 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (2.9.0.post0)
Requirement already satisfied: google-auth>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (2.38.0)
Requirement already satisfied: websocket-client!=0.40.0,!=0.41.*,!=0.42.*,>=0.32.0 in /usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0->chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4.0->langchain-chroma) (1.8.0)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-

packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.32.3)
Requirement already satisfied: requests-oauthlib in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.0.0)
Requirement already satisfied: oauthlib>=3.2.2 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.2.2)
Requirement already satisfied: urllib3>=1.24.2 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.4.0)
Requirement already satisfied: durationpy>=0.7 in
/usr/local/lib/python3.11/dist-packages (from kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.9)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-core>=0.3.52->langchain-chroma) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-core>=0.3.52->langchain-chroma) (0.23.0)
Requirement already satisfied: coloredlogs in /usr/local/lib/python3.11/dist-
packages (from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (15.0.1)
Requirement already satisfied: flatbuffers in /usr/local/lib/python3.11/dist-
packages (from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (25.2.10)
Requirement already satisfied: protobuf in /usr/local/lib/python3.11/dist-
packages (from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (5.29.4)
Requirement already satisfied: sympy in /usr/local/lib/python3.11/dist-packages
(from onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.13.1)
Requirement already satisfied: deprecated>=1.2.6 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-api>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.2.18)
Requirement already satisfied: importlib-metadata<8.7.0,>=6.0 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-api>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4

```
.0->langchain-chroma) (8.6.1)
Requirement already satisfied: googleapis-common-protos~=1.52 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.70.0)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-common==1.32.1
in /usr/local/lib/python3.11/dist-packages (from opentelemetry-exporter-otlp-
proto-grpc>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.32.1)
Requirement already satisfied: opentelemetry-proto==1.32.1 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-exporter-otlp-proto-
grpc>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.32.1)
Requirement already satisfied: opentelemetry-instrumentation-asgi==0.53b1 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-instrumentation-fast
api>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.53b1)
Requirement already satisfied: opentelemetry-instrumentation==0.53b1 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-instrumentation-fast
api>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.53b1)
Requirement already satisfied: opentelemetry-semantic-conventions==0.53b1 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-instrumentation-fast
api>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.53b1)
Requirement already satisfied: opentelemetry-util-http==0.53b1 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-instrumentation-fast
api>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.53b1)
Requirement already satisfied: wrapt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from opentelemetry-
instrumentation==0.53b1->opentelemetry-instrumentation-fastapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.17.2)
Requirement already satisfied: asgiref~=3.0 in /usr/local/lib/python3.11/dist-
packages (from opentelemetry-instrumentation-asgi==0.53b1->opentelemetry-
instrumentation-fastapi>=0.41b0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.8.1)
Requirement already satisfied: backoff>=1.10.0 in
/usr/local/lib/python3.11/dist-packages (from posthog>=2.4.0-
```

```
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.2.1)
Requirement already satisfied: distro>=1.5.0 in /usr/lib/python3/dist-packages
(from posthog>=2.4.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.7.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=1.9-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=1.9-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.33.1)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=1.9-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.4.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.11/dist-packages (from rich>=10.11.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/dist-packages (from rich>=10.11.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2.19.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in
/usr/local/lib/python3.11/dist-packages (from tokenizers>=0.13.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.30.2)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.11/dist-
packages (from typer>=0.9.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (8.1.8)
Requirement already satisfied: shellingham>=1.3.0 in
/usr/local/lib/python3.11/dist-packages (from typer>=0.9.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.5.4)
Requirement already satisfied: httptools>=0.6.3 in
/usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.6.4)
Requirement already satisfied: python-dotenv>=0.13 in
/usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.1.0)
Requirement already satisfied: uvloop!=0.15.0,!=0.15.1,>=0.14.0 in
/usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3-
```

>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.21.0)
Requirement already satisfied: watchfiles>=0.13 in
/usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.0.5)
Requirement already satisfied: websockets>=10.4 in
/usr/local/lib/python3.11/dist-packages (from uvicorn[standard]>=0.18.3-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (15.0.1)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from google-auth>=1.0.1-
>kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (5.5.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.11/dist-packages (from google-auth>=1.0.1-
>kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.4.2)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-
packages (from google-auth>=1.0.1->kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (4.9.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-
packages (from huggingface-hub<1.0,>=0.16.4->tokenizers>=0.13.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.18.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.16.4-
>tokenizers>=0.13.2-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (2025.3.2)
Requirement already satisfied: zipp>=3.20 in /usr/local/lib/python3.11/dist-
packages (from importlib-metadata<8.7.0,>=6.0->opentelemetry-api>=1.2.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.21.0)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-
packages (from markdown-it-py>=2.2.0->rich>=10.11.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (3.4.1)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx>=0.27.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4

```
.0->langchain-chroma) (1.3.1)
Requirement already satisfied: humanfriendly>=9.1 in
/usr/local/lib/python3.11/dist-packages (from coloredlogs->onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (10.0)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy->onnxruntime>=1.14.1-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (1.3.0)
Requirement already satisfied: pyasn1<0.7.0,>=0.6.1 in
/usr/local/lib/python3.11/dist-packages (from pyasn1-modules>=0.2.1->google-auth
>=1.0.1->kubernetes>=28.1.0-
>chromadb!=0.5.10,!=0.5.11,!=0.5.12,!=0.5.4,!=0.5.5,!=0.5.7,!=0.5.9,<0.7.0,>=0.4
.0->langchain-chroma) (0.6.1)
Collecting langchain-community
  Using cached langchain_community-0.3.23-py3-none-any.whl.metadata (2.5 kB)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.56 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.58)
Requirement already satisfied: langchain<1.0.0,>=0.3.24 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.25)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (2.0.40)
Requirement already satisfied: requests<3,>=2 in /usr/local/lib/python3.11/dist-
packages (from langchain-community) (2.32.3)
Requirement already satisfied: PyYAML>=5.3 in /usr/local/lib/python3.11/dist-
packages (from langchain-community) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (3.11.15)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (9.1.2)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain-community)
  Using cached dataclasses_json-0.6.7-py3-none-any.whl.metadata (25 kB)
Collecting pydantic-settings<3.0.0,>=2.4.0 (from langchain-community)
  Using cached pydantic_settings-2.9.1-py3-none-any.whl.metadata (3.8 kB)
Requirement already satisfied: langsmith<0.4,>=0.1.125 in
/usr/local/lib/python3.11/dist-packages (from langchain-community) (0.3.41)
Collecting httpx-sse<1.0.0,>=0.4.0 (from langchain-community)
  Using cached httpx_sse-0.4.0-py3-none-any.whl.metadata (9.0 kB)
Requirement already satisfied: numpy>=1.26.2 in /usr/local/lib/python3.11/dist-
packages (from langchain-community) (2.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-
packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (25.3.0)
```

```
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (6.4.3)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.11/dist-packages (from aiohttp<4.0.0,>=3.8.3->langchain-
community) (1.20.0)
Collecting marshmallow<4.0.0,>=3.18.0 (from dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Using cached marshmallow-3.26.1-py3-none-any.whl.metadata (7.3 kB)
Collecting typing-inspect<1,>=0.4.0 (from dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Using cached typing_inspect-0.9.0-py3-none-any.whl.metadata (1.5 kB)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.8 in
/usr/local/lib/python3.11/dist-packages (from
langchain<1.0.0,>=0.3.24->langchain-community) (0.3.8)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in
/usr/local/lib/python3.11/dist-packages (from
langchain<1.0.0,>=0.3.24->langchain-community) (2.11.3)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.56->langchain-community) (1.33)
Requirement already satisfied: packaging<25,>=23.2 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.56->langchain-community) (24.2)
Requirement already satisfied: typing-extensions>=4.7 in
/usr/local/lib/python3.11/dist-packages (from langchain-
core<1.0.0,>=0.3.56->langchain-community) (4.13.2)
Requirement already satisfied: httpx<1,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (0.28.1)
Requirement already satisfied: orjson<4.0.0,>=3.9.14 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (3.10.18)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in
/usr/local/lib/python3.11/dist-packages (from
langsmith<0.4,>=0.1.125->langchain-community) (0.23.0)
Requirement already satisfied: python-dotenv>=0.21.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic-
settings<3.0.0,>=2.4.0->langchain-community) (1.1.0)
```

Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic-
settings<3.0.0,>=2.4.0->langchain-community) (0.4.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-
community) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests<3,>=2->langchain-community) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-
community) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests<3,>=2->langchain-
community) (2025.4.26)
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-
packages (from SQLAlchemy<3,>=1.4->langchain-community) (3.2.1)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community) (4.9.0)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community)
(1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from
httpcore==1.*->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-community)
(0.16.0)
Requirement already satisfied: jsonpointer>=1.9 in
/usr/local/lib/python3.11/dist-packages (from jsonpatch<2.0,>=1.33->langchain-
core<1.0.0,>=0.3.56->langchain-community) (3.0.0)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from
pydantic<3.0.0,>=2.7.4->langchain<1.0.0,>=0.3.24->langchain-community) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in
/usr/local/lib/python3.11/dist-packages (from
pydantic<3.0.0,>=2.7.4->langchain<1.0.0,>=0.3.24->langchain-community) (2.33.1)
Collecting mypy-extensions>=0.3.0 (from typing-inspect<1,>=0.4.0->dataclasses-
json<0.7,>=0.5.7->langchain-community)
  Using cached mypy_extensions-1.1.0-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-
packages (from anyio->httpx<1,>=0.23.0->langsmith<0.4,>=0.1.125->langchain-
community) (1.3.1)
Using cached langchain_community-0.3.23-py3-none-any.whl (2.5 MB)
Using cached dataclasses_json-0.6.7-py3-none-any.whl (28 kB)
Using cached httpx_sse-0.4.0-py3-none-any.whl (7.8 kB)
Using cached pydantic_settings-2.9.1-py3-none-any.whl (44 kB)
Using cached marshmallow-3.26.1-py3-none-any.whl (50 kB)
Using cached typing_inspect-0.9.0-py3-none-any.whl (8.8 kB)
Using cached mypy_extensions-1.1.0-py3-none-any.whl (5.0 kB)
Installing collected packages: mypy-extensions, marshmallow, httpx-sse, typing-

```
inspect, pydantic-settings, dataclasses-json, langchain-community
Successfully installed dataclasses-json-0.6.7 httpx-sse-0.4.0 langchain-
community-0.3.23 marshmallow-3.26.1 mypy-extensions-1.1.0 pydantic-
settings-2.9.1 typing-inspect-0.9.0
Collecting tiktoken
  Downloading tiktoken-0.9.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.7 kB)
Requirement already satisfied: regex>=2022.1.18 in
/usr/local/lib/python3.11/dist-packages (from tiktoken) (2024.11.6)
Requirement already satisfied: requests>=2.26.0 in
/usr/local/lib/python3.11/dist-packages (from tiktoken) (2.32.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken)
(3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests>=2.26.0->tiktoken) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken)
(2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests>=2.26.0->tiktoken)
(2025.4.26)
Downloading
tiktoken-0.9.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2
MB)
                         1.2/1.2 MB
13.2 MB/s eta 0:00:00
Installing collected packages: tiktoken
Successfully installed tiktoken-0.9.0
Collecting unstructured
  Downloading unstructured-0.17.2-py3-none-any.whl.metadata (24 kB)
Collecting chardet (from unstructured)
  Downloading chardet-5.2.0-py3-none-any.whl.metadata (3.4 kB)
Collecting filetype (from unstructured)
  Downloading filetype-1.2.0-py2.py3-none-any.whl.metadata (6.5 kB)
Collecting python-magic (from unstructured)
  Downloading python_magic-0.4.27-py2.py3-none-any.whl.metadata (5.8 kB)
Collecting lxml (from unstructured)
  Downloading lxml-5.4.0-cp311-cp311-manylinux_2_28_x86_64.whl.metadata (3.5 kB)
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages
(from unstructured) (3.9.1)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-
packages (from unstructured) (2.32.3)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.11/dist-
packages (from unstructured) (4.13.4)
Collecting emoji (from unstructured)
  Downloading emoji-2.14.1-py3-none-any.whl.metadata (5.7 kB)
Requirement already satisfied: dataclasses-json in
```

/usr/local/lib/python3.11/dist-packages (from unstructured) (0.6.7)
Collecting python-iso639 (from unstructured)
  Downloading python_iso639-2025.2.18-py3-none-any.whl.metadata (14 kB)
Collecting langdetect (from unstructured)
  Downloading langdetect-1.0.9.tar.gz (981 kB)
                           981.5/981.5 kB
16.9 MB/s eta 0:00:00
  Installing build dependencies … done
  Getting requirements to build wheel … done
  Preparing metadata (pyproject.toml) … done
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages
(from unstructured) (2.0.2)
Collecting rapidfuzz (from unstructured)
  Downloading rapidfuzz-3.13.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Requirement already satisfied: backoff in /usr/local/lib/python3.11/dist-
packages (from unstructured) (2.2.1)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.11/dist-packages (from unstructured) (4.13.2)
Collecting unstructured-client (from unstructured)
  Downloading unstructured_client-0.34.0-py3-none-any.whl.metadata (21 kB)
Requirement already satisfied: wrapt in /usr/local/lib/python3.11/dist-packages
(from unstructured) (1.17.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages
(from unstructured) (4.67.1)
Requirement already satisfied: psutil in /usr/local/lib/python3.11/dist-packages
(from unstructured) (5.9.5)
Collecting python-oxmsg (from unstructured)
  Downloading python_oxmsg-0.0.2-py3-none-any.whl.metadata (5.0 kB)
Collecting html5lib (from unstructured)
  Downloading html5lib-1.1-py2.py3-none-any.whl.metadata (16 kB)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.11/dist-
packages (from beautifulsoup4->unstructured) (2.7)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in
/usr/local/lib/python3.11/dist-packages (from dataclasses-json->unstructured)
(3.26.1)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from dataclasses-json->unstructured)
(0.9.0)
Requirement already satisfied: six>=1.9 in /usr/local/lib/python3.11/dist-
packages (from html5lib->unstructured) (1.17.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.11/dist-
packages (from html5lib->unstructured) (0.5.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages
(from nltk->unstructured) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages
(from nltk->unstructured) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in

/usr/local/lib/python3.11/dist-packages (from nltk->unstructured) (2024.11.6)
Collecting olefile (from python-oxmsg->unstructured)
  Downloading olefile-0.47-py2.py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->unstructured) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
packages (from requests->unstructured) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.11/dist-packages (from requests->unstructured) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.11/dist-packages (from requests->unstructured)
(2025.4.26)
Collecting aiofiles>=24.1.0 (from unstructured-client->unstructured)
  Downloading aiofiles-24.1.0-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: cryptography>=3.1 in /usr/lib/python3/dist-
packages (from unstructured-client->unstructured) (3.4.8)
Collecting eval-type-backport>=0.2.0 (from unstructured-client->unstructured)
  Downloading eval_type_backport-0.2.2-py3-none-any.whl.metadata (2.2 kB)
Requirement already satisfied: httpx>=0.27.0 in /usr/local/lib/python3.11/dist-
packages (from unstructured-client->unstructured) (0.28.1)
Requirement already satisfied: nest-asyncio>=1.6.0 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(1.6.0)
Requirement already satisfied: pydantic>=2.11.2 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(2.11.3)
Collecting pypdf>=4.0 (from unstructured-client->unstructured)
  Downloading pypdf-5.4.0-py3-none-any.whl.metadata (7.3 kB)
Requirement already satisfied: requests-toolbelt>=1.0.0 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(1.0.0)
Requirement already satisfied: typing-inspection>=0.4.0 in
/usr/local/lib/python3.11/dist-packages (from unstructured-client->unstructured)
(0.4.0)
Requirement already satisfied: anyio in /usr/local/lib/python3.11/dist-packages
(from httpx>=0.27.0->unstructured-client->unstructured) (4.9.0)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.11/dist-
packages (from httpx>=0.27.0->unstructured-client->unstructured) (1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.11/dist-
packages (from httpcore==1.*->httpx>=0.27.0->unstructured-client->unstructured)
(0.16.0)
Requirement already satisfied: packaging>=17.0 in
/usr/local/lib/python3.11/dist-packages (from
marshmallow<4.0.0,>=3.18.0->dataclasses-json->unstructured) (24.2)
Requirement already satisfied: annotated-types>=0.6.0 in
/usr/local/lib/python3.11/dist-packages (from pydantic>=2.11.2->unstructured-
client->unstructured) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.1 in

/usr/local/lib/python3.11/dist-packages (from pydantic>=2.11.2->unstructured-client->unstructured) (2.33.1)
Requirement already satisfied: mypy-extensions>=0.3.0 in
/usr/local/lib/python3.11/dist-packages (from typing-inspect<1,>=0.4.0->dataclasses-json->unstructured) (1.1.0)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.11/dist-packages (from anyio->httpx>=0.27.0->unstructured-client->unstructured) (1.3.1)
Downloading unstructured-0.17.2-py3-none-any.whl (1.8 MB)
                          1.8/1.8 MB
53.1 MB/s eta 0:00:00
Downloading chardet-5.2.0-py3-none-any.whl (199 kB)
                          199.4/199.4 kB
14.0 MB/s eta 0:00:00
Downloading emoji-2.14.1-py3-none-any.whl (590 kB)
                          590.6/590.6 kB
30.8 MB/s eta 0:00:00
Downloading filetype-1.2.0-py2.py3-none-any.whl (19 kB)
Downloading html5lib-1.1-py2.py3-none-any.whl (112 kB)
                          112.2/112.2 kB
7.5 MB/s eta 0:00:00
Downloading lxml-5.4.0-cp311-cp311-manylinux_2_28_x86_64.whl (4.9 MB)
                          4.9/4.9 MB
92.4 MB/s eta 0:00:00
Downloading python_iso639-2025.2.18-py3-none-any.whl (167 kB)
                          167.6/167.6 kB
12.3 MB/s eta 0:00:00
Downloading python_magic-0.4.27-py2.py3-none-any.whl (13 kB)
Downloading python_oxmsg-0.0.2-py3-none-any.whl (31 kB)
Downloading
rapidfuzz-3.13.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.1 MB)
                          3.1/3.1 MB
86.6 MB/s eta 0:00:00
Downloading unstructured_client-0.34.0-py3-none-any.whl (189 kB)
                          189.4/189.4 kB
13.3 MB/s eta 0:00:00
Downloading aiofiles-24.1.0-py3-none-any.whl (15 kB)
Downloading eval_type_backport-0.2.2-py3-none-any.whl (5.8 kB)
Downloading pypdf-5.4.0-py3-none-any.whl (302 kB)
                          302.3/302.3 kB
16.8 MB/s eta 0:00:00
Downloading olefile-0.47-py2.py3-none-any.whl (114 kB)
                          114.6/114.6 kB
7.6 MB/s eta 0:00:00
Building wheels for collected packages: langdetect
  Building wheel for langdetect (pyproject.toml) … done
  Created wheel for langdetect: filename=langdetect-1.0.9-py3-none-any.whl
size=993331

16

sha256=e3857d45451b2e1d982ca9ca1a12cbb500b92914477c5b3932b4ba8e7b583811
    Stored in directory: /root/.cache/pip/wheels/0a/f2/b2/e5ca405801e05eb7c8ed5b3b
4bcf1fcabcd6272c167640072e
Successfully built langdetect
Installing collected packages: filetype, rapidfuzz, python-magic, python-iso639,
pypdf, olefile, lxml, langdetect, html5lib, eval-type-backport, emoji, chardet,
aiofiles, python-oxmsg, unstructured-client, unstructured
Successfully installed aiofiles-24.1.0 chardet-5.2.0 emoji-2.14.1 eval-type-
backport-0.2.2 filetype-1.2.0 html5lib-1.1 langdetect-1.0.9 lxml-5.4.0
olefile-0.47 pypdf-5.4.0 python-iso639-2025.2.18 python-magic-0.4.27 python-
oxmsg-0.0.2 rapidfuzz-3.13.0 unstructured-0.17.2 unstructured-client-0.34.0

```
[ ]: !pip install "tensorflow>=2.15.0"
     !pip install --upgrade tensorflow-hub
```

Collecting tensorflow>=2.15.0
  Downloading tensorflow-2.19.0-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (4.1 kB)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.11/dist-
packages (from tensorflow>=2.15.0) (1.4.0)
Collecting astunparse>=1.6.0 (from tensorflow>=2.15.0)
  Downloading astunparse-1.6.3-py2.py3-none-any.whl.metadata (4.4 kB)
Requirement already satisfied: flatbuffers>=24.3.25 in
/usr/local/lib/python3.11/dist-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
/usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (0.6.0)
Collecting google-pasta>=0.1.1 (from tensorflow>=2.15.0)
  Downloading google_pasta-0.2.0-py3-none-any.whl.metadata (814 bytes)
Collecting libclang>=13.0.0 (from tensorflow>=2.15.0)
  Downloading libclang-18.1.1-py2.py3-none-manylinux2010_x86_64.whl.metadata
(5.2 kB)
Requirement already satisfied: opt-einsum>=2.3.2 in
/usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (3.4.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-
packages (from tensorflow>=2.15.0) (24.2)
Requirement already satisfied:
protobuf!=4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<6.0.0dev,>=3.20.3
in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (5.29.4)
Requirement already satisfied: requests<3,>=2.21.0 in
/usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (2.32.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-
packages (from tensorflow>=2.15.0) (75.2.0)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.11/dist-
packages (from tensorflow>=2.15.0) (1.17.0)
Requirement already satisfied: termcolor>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (3.0.1)
Requirement already satisfied: typing-extensions>=3.6.6 in
/usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (4.13.2)

Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (1.17.2)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (1.71.0)
Collecting tensorboard~=2.19.0 (from tensorflow>=2.15.0)
  Downloading tensorboard-2.19.0-py3-none-any.whl.metadata (1.8 kB)
Requirement already satisfied: keras>=3.5.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (3.8.0)
Requirement already satisfied: numpy<2.2.0,>=1.26.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (2.0.2)
Requirement already satisfied: h5py>=3.11.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (3.13.0)
Requirement already satisfied: ml-dtypes<1.0.0,>=0.5.1 in /usr/local/lib/python3.11/dist-packages (from tensorflow>=2.15.0) (0.5.1)
Collecting tensorflow-io-gcs-filesystem>=0.23.1 (from tensorflow>=2.15.0)
  Downloading tensorflow_io_gcs_filesystem-0.37.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (14 kB)
Collecting wheel<1.0,>=0.23.0 (from astunparse>=1.6.0->tensorflow>=2.15.0)
  Downloading wheel-0.45.1-py3-none-any.whl.metadata (2.3 kB)
Requirement already satisfied: rich in /usr/local/lib/python3.11/dist-packages (from keras>=3.5.0->tensorflow>=2.15.0) (14.0.0)
Requirement already satisfied: namex in /usr/local/lib/python3.11/dist-packages (from keras>=3.5.0->tensorflow>=2.15.0) (0.0.9)
Requirement already satisfied: optree in /usr/local/lib/python3.11/dist-packages (from keras>=3.5.0->tensorflow>=2.15.0) (0.15.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow>=2.15.0) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow>=2.15.0) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow>=2.15.0) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow>=2.15.0) (2025.4.26)
Requirement already satisfied: markdown>=2.6.8 in /usr/lib/python3/dist-packages (from tensorboard~=2.19.0->tensorflow>=2.15.0) (3.3.6)
Collecting tensorboard-data-server<0.8.0,>=0.7.0 (from tensorboard~=2.19.0->tensorflow>=2.15.0)
  Downloading tensorboard_data_server-0.7.2-py3-none-manylinux_2_31_x86_64.whl.metadata (1.1 kB)
Collecting werkzeug>=1.0.1 (from tensorboard~=2.19.0->tensorflow>=2.15.0)
  Downloading werkzeug-3.1.3-py3-none-any.whl.metadata (3.7 kB)
Requirement already satisfied: MarkupSafe>=2.1.1 in /usr/local/lib/python3.11/dist-packages (from werkzeug>=1.0.1->tensorboard~=2.19.0->tensorflow>=2.15.0) (3.0.2)
Requirement already satisfied: markdown-it-py>=2.2.0 in

```
/usr/local/lib/python3.11/dist-packages (from
rich->keras>=3.5.0->tensorflow>=2.15.0) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/dist-packages (from
rich->keras>=3.5.0->tensorflow>=2.15.0) (2.19.1)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-
packages (from markdown-it-py>=2.2.0->rich->keras>=3.5.0->tensorflow>=2.15.0)
(0.1.2)
Downloading
tensorflow-2.19.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(644.9 MB)
                          644.9/644.9 MB
1.6 MB/s eta 0:00:00
Downloading astunparse-1.6.3-py2.py3-none-any.whl (12 kB)
Downloading google_pasta-0.2.0-py3-none-any.whl (57 kB)
                          57.5/57.5 kB
3.9 MB/s eta 0:00:00
Downloading libclang-18.1.1-py2.py3-none-manylinux2010_x86_64.whl (24.5
MB)
                          24.5/24.5 MB
82.1 MB/s eta 0:00:00
Downloading tensorboard-2.19.0-py3-none-any.whl (5.5 MB)
                          5.5/5.5 MB
102.1 MB/s eta 0:00:00
Downloading tensorflow_io_gcs_filesystem-0.37.1-cp311-cp311-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (5.1 MB)
                          5.1/5.1 MB
105.6 MB/s eta 0:00:00
Downloading tensorboard_data_server-0.7.2-py3-none-
manylinux_2_31_x86_64.whl (6.6 MB)
                          6.6/6.6 MB
116.2 MB/s eta 0:00:00
Downloading werkzeug-3.1.3-py3-none-any.whl (224 kB)
                          224.5/224.5 kB
13.1 MB/s eta 0:00:00
Downloading wheel-0.45.1-py3-none-any.whl (72 kB)
                          72.5/72.5 kB
4.8 MB/s eta 0:00:00
Installing collected packages: libclang, wheel, werkzeug, tensorflow-io-
gcs-filesystem, tensorboard-data-server, google-pasta, tensorboard, astunparse,
tensorflow
Successfully installed astunparse-1.6.3 google-pasta-0.2.0 libclang-18.1.1
tensorboard-2.19.0 tensorboard-data-server-0.7.2 tensorflow-2.19.0 tensorflow-
io-gcs-filesystem-0.37.1 werkzeug-3.1.3 wheel-0.45.1
Collecting tensorflow-hub
  Downloading tensorflow_hub-0.16.1-py2.py3-none-any.whl.metadata (1.3 kB)
Requirement already satisfied: numpy>=1.12.0 in /usr/local/lib/python3.11/dist-
packages (from tensorflow-hub) (2.0.2)
```

```
Requirement already satisfied: protobuf>=3.19.6 in
/usr/local/lib/python3.11/dist-packages (from tensorflow-hub) (5.29.4)
Collecting tf-keras>=2.14.1 (from tensorflow-hub)
  Downloading tf_keras-2.19.0-py3-none-any.whl.metadata (1.8 kB)
Requirement already satisfied: tensorflow<2.20,>=2.19 in
/usr/local/lib/python3.11/dist-packages (from tf-keras>=2.14.1->tensorflow-hub)
(2.19.0)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.11/dist-
packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (1.6.3)
Requirement already satisfied: flatbuffers>=24.3.25 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (25.2.10)
Requirement already satisfied: gast!=0.5.0,!=0.5.1,!=0.5.2,>=0.2.1 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (0.6.0)
Requirement already satisfied: google-pasta>=0.1.1 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (0.2.0)
Requirement already satisfied: libclang>=13.0.0 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (18.1.1)
Requirement already satisfied: opt-einsum>=2.3.2 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (3.4.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-
packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (24.2)
Requirement already satisfied: requests<3,>=2.21.0 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (2.32.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-
packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(75.2.0)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.11/dist-
packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(1.17.0)
Requirement already satisfied: termcolor>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (3.0.1)
Requirement already satisfied: typing-extensions>=3.6.6 in
/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (4.13.2)
Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.11/dist-
packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(1.17.2)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in
```

/usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (1.71.0)
Requirement already satisfied: tensorboard~=2.19.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (2.19.0)
Requirement already satisfied: keras>=3.5.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (3.8.0)
Requirement already satisfied: h5py>=3.11.0 in /usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (3.13.0)
Requirement already satisfied: ml-dtypes<1.0.0,>=0.5.1 in /usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (0.5.1)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.11/dist-packages (from tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (0.37.1)
Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.11/dist-packages (from astunparse>=1.6.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (0.45.1)
Requirement already satisfied: rich in /usr/local/lib/python3.11/dist-packages (from keras>=3.5.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (14.0.0)
Requirement already satisfied: namex in /usr/local/lib/python3.11/dist-packages (from keras>=3.5.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (0.0.9)
Requirement already satisfied: optree in /usr/local/lib/python3.11/dist-packages (from keras>=3.5.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (0.15.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.21.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (2025.4.26)
Requirement already satisfied: markdown>=2.6.8 in /usr/lib/python3/dist-packages (from tensorboard~=2.19.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub) (3.3.6)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in

```
/usr/local/lib/python3.11/dist-packages (from
tensorboard~=2.19.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(0.7.2)
Requirement already satisfied: werkzeug>=1.0.1 in
/usr/local/lib/python3.11/dist-packages (from
tensorboard~=2.19.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(3.1.3)
Requirement already satisfied: MarkupSafe>=2.1.1 in
/usr/local/lib/python3.11/dist-packages (from
werkzeug>=1.0.1->tensorboard~=2.19.0->tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (3.0.2)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.11/dist-packages (from
rich->keras>=3.5.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/dist-packages (from
rich->keras>=3.5.0->tensorflow<2.20,>=2.19->tf-keras>=2.14.1->tensorflow-hub)
(2.19.1)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-
packages (from markdown-it-
py>=2.2.0->rich->keras>=3.5.0->tensorflow<2.20,>=2.19->tf-
keras>=2.14.1->tensorflow-hub) (0.1.2)
Downloading tensorflow_hub-0.16.1-py2.py3-none-any.whl (30 kB)
Downloading tf_keras-2.19.0-py3-none-any.whl (1.7 MB)
                        1.7/1.7 MB
24.4 MB/s eta 0:00:00
Installing collected packages: tf-keras, tensorflow-hub
Successfully installed tensorflow-hub-0.16.1 tf-keras-2.19.0
```

```python
[ ]:
```

```python
[1]: # Import the libraries

import os
import sys
from IPython.display import HTML, display

import numpy as np
import pandas as pd
import tensorflow as tf
from math import ceil
from scipy.spatial.distance import cosine

import matplotlib.pyplot as plt
import seaborn as sns
```

```
import collections
import random
import time
import string
import re

import nltk
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('punkt_tab')
nltk.download('wordnet')
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from nltk.tokenize import word_tokenize
from nltk.tokenize import sent_tokenize

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
from sklearn.metrics import roc_curve, auc
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report

from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.models import Sequential, Model
from tensorflow.keras.layers import Input, Embedding, Dense, Convolution1D,␣
 ↪MaxPooling1D, GlobalMaxPooling1D, Flatten, Dropout, LSTM, Bidirectional
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data…
[nltk_data]    Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data…
[nltk_data]    Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package punkt_tab to /root/nltk_data…
[nltk_data]    Unzipping tokenizers/punkt_tab.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data…
```

```
[2]: from google.colab import drive
     drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
[3]: %cd /content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app
```

```
/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app
```

```
[4]:  !ls
```

```
df.csv                      MutualFundSummary10
df_main.csv                 MutualFundSummary2
df_RAG2_5.csv               MutualFundSummary2_5
df_RAG2.csv                 MutualFundSummary3
df_RAG3_5.csv               MutualFundSummary3_5
df_RAG3.csv                 MutualFundSummary4
df_RAG4.csv                 MutualFundSummary4_5
df_RAG_all.csv              MutualFundSummary5
df_RAG.csv                  MutualFundSummary6
df_summary.csv              MutualFundSummary7
docs                        MutualFundSummary8
FundArticlesHealthy.csv     MutualFundSummary9
FundArticlesUnhealthy.csv   NLP_Deep_Learning_Questrom_WS_03_2020.ipynb
glove.6B.50d.txt            word2vec_model.pkl
MutualFundLabels.csv        word2vec_MutualFundSummary.txt
MutualFundSummary           word2vec_perso.txt
MutualFundSummary1
```

```
[5]:  DIRECTORY = '/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app'

      SUMMARY_PATH = '/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/NLP_app/
       ↪MutualFundSummary'
      SUMMARY_LABELS_PATH = '/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/
       ↪NLP_app/MutualFundLabels.csv'

      HEALTHY_ARTICLES_PATH = os.path.join(DIRECTORY, "Data", "FundArticlesHealthy.
       ↪csv")
      UNHEALTHY_ARTICLES_PATH = os.path.join(DIRECTORY, "Data",␣
       ↪"FundArticlesUnhealthy.csv")

      glove_word2vec = 'glove.6B.50d.txt'
      our_word2vec = 'word2vec_perso.txt'
      new_word2vec = 'word2vec_MutualFundSummary.txt'
```

```
[6]:  # Progress bar
      def progress(value, max=100):
          return HTML("""
              <progress
                  value='{value}'
                  max='{max}',
                  style='width: 100%'
              >
                  {value}
              </progress>
          """.format(value=value, max=max))
```

```python
# Save a word2vec dictionary.
def save_word2vec(filename):
    #with open(os.path.join('/Users/haoxing/Documents/Work/Teaching/Machine
 ↪learning for Finance/Codes/NLP/NLP_app', filename),'a' , encoding='utf-8')
 ↪as f :
    with open(os.path.join('/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/
 ↪NLP_app/', filename),'a' , encoding='utf-8') as f :
        for k, v in word2vec.items():
            line = k+' '+str(list(v)).strip('[]').replace(',','')+'\n'
            f.write(line)


# Load a word2vec dictionary.
def load_word2vec(filename):
    word2vec = {}
    #with open(os.path.join('/Users/haoxing/Documents/Work/Teaching/Machine
 ↪learning for Finance/Codes/NLP/NLP_app', filename), encoding='utf8') as f:
    with open(os.path.join('/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/
 ↪NLP_app/', filename), encoding='utf8') as f:
        for line in f:
            try :
                values = line.split()
                word = values[0]
                vec = np.asarray(values[1:], dtype='float32')
                word2vec[word] = vec
            except :
                None
    return word2vec


# read the repo in PATH and append the texts in a list
def get_data(PATH):
    list_dir = os.listdir(PATH)
    texts = []
    fund_names = []
    out = display(progress(0, len(list_dir)-1), display_id=True)
    for ii, filename in enumerate(list_dir) :
        with open(PATH+'/'+filename, 'r', encoding="utf8") as f :
            txt = f.read()
            try :
                txt_split = txt.split('<head_breaker>')
                summary = txt_split[1].strip()
                fund_name = txt_split[0].strip()
            except :
                summary = txt
                fund_name = ''
        texts.append(summary)
        fund_names.append(fund_name)
```

```
           out.update(progress(ii, len(list_dir)-1))
       return fund_names, texts
```

[7]:
```
# Get the summaries
fund_names, summaries = get_data(SUMMARY_PATH)
```

<IPython.core.display.HTML object>

[8]:
```
len(fund_names)
```

[8]: 545

[9]:
```
len(summaries)
```

[9]: 545

[10]:
```
fund_names[20]
```

[10]: 'AB Municipal Bond Inflation Strategy'

[11]:
```
type(summaries), type(fund_names)
```

[11]: (list, list)

[12]:
```
df_summary = pd.read_csv('/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/
 ↪NLP_app/df_summary.csv')
df_summary
```

[12]:
```
                     id                                      fund_name  \
0    0000051931-18-000151                American Funds College 2018 Fund
1    0000051931-18-000151                American Funds College 2021 Fund
2    0000051931-18-000151                American Funds College 2024 Fund
3    0000051931-18-000151                American Funds College 2027 Fund
4    0000051931-18-000151                American Funds College 2030 Fund
..                    ...                                            ...
427  0001710607-18-000172    American Century Diversified Corporate Bond ETF
428  0001710607-18-000172    American Century Diversified Municipal Bond ETF
429  0001710607-18-000172  American Century Quality Diversified Internati…
430  0001710607-18-000172      American Century STOXX U.S. Quality Growth ETF
431  0001710607-18-000172       American Century STOXX U.S. Quality Value ETF

    Performance fee?            Ivestment Strategy Leverage?  \
0                NaN        Balanced Fund (Low Risk)       Yes
1                NaN        Balanced Fund (Low Risk)       Yes
2                NaN        Balanced Fund (Low Risk)       Yes
3                NaN        Balanced Fund (Low Risk)       Yes
4                NaN        Balanced Fund (Low Risk)       Yes
```

26

..	…	…	…
427	NaN	Fixed Income Long Only (Low Risk)	Yes
428	NaN	Fixed Income Long Only (Low Risk)	No
429	NaN	Equity Long Only (Low Risk)	Yes
430	NaN	Equity Long Only (Low Risk)	Yes
431	NaN	Equity Long Only (Low Risk)	Yes

	Portfolio composition	Concentration \
0	Investment grade securities	Diversified
1	Investment grade securities	Diversified
2	Investment grade securities	Diversified
3	Investment grade securities	Diversified
4	Investment grade securities	Diversified
..	…	…
427	Investment grade securities	Diversified
428	Investment grade securities	Diversified
429	Sub-investment grade securities or emerging ma…	Diversified
430	Listed Equities	Diversified
431	Listed Equities	Diversified

	Ivestment_Strategy_label	Predicted investment strategy \
0	2.0	Balanced Fund (Low Risk)
1	2.0	Balanced Fund (Low Risk)
2	2.0	Balanced Fund (Low Risk)
3	2.0	Balanced Fund (Low Risk)
4	2.0	Balanced Fund (Low Risk)
..	…	…
427	1.0	Fixed Income Long Only (Low Risk)
428	1.0	Fixed Income Long Only (Low Risk)
429	0.0	Equity Long Only (Low Risk)
430	0.0	Equity Long Only (Low Risk)
431	0.0	Equity Long Only (Low Risk)

	investment strategy evidence \
0	American Funds College 2018 Fund <head_breaker…
1	American Funds College 2021 Fund <head_breaker…
2	American Funds College 2024 Fund <head_breaker…
3	American Funds College 2027 Fund <head_breaker…
4	American Funds College 2030 Fund <head_breaker…
..	…
427	American Century Diversified Corporate Bond ET…
428	American Century Diversified Municipal Bond ET…
429	American Century Quality Diversified Internati…
430	American Century STOXX U.S. Quality Growth ETF…
431	American Century STOXX U.S. Quality Value ETF …

source meta data Prediction

```
0     {'source': 'MutualFundSummary/American Funds C…        2.0
1     {'source': 'MutualFundSummary/American Funds C…        2.0
2     {'source': 'MutualFundSummary1/American Funds …        2.0
3     {'source': 'MutualFundSummary1/American Funds …        2.0
4     {'source': 'MutualFundSummary/American Funds C…        2.0
..                                               …          …
427   {'source': 'MutualFundSummary/American Century…        1.0
428   {'source': 'MutualFundSummary/American Century…        1.0
429   {'source': 'MutualFundSummary/American Century…        0.0
430   {'source': 'MutualFundSummary/American Century…        0.0
431   {'source': 'MutualFundSummary/American Century…        0.0

[432 rows x 12 columns]
```

```python
[13]: common_fund_names = []
      for index, name in enumerate(fund_names):
          # Check if the current fund name is in the set of common names
          if name in set(df_summary['fund_name']):
              # If it is, append the corresponding summary (using the same index)
              common_fund_names.append(fund_names[index])
      len(common_fund_names)
```

```
[13]: 432
```

```python
[14]: common_summaries = []
      for index, name in enumerate(fund_names):
          # Check if the current fund name is in the set of common names
          if name in set(df_summary['fund_name']):
              # If it is, append the corresponding summary (using the same index)
              common_summaries.append(summaries[index])
      len(common_summaries)
```

```
[14]: 432
```

```python
[15]: #checking
      common_fund_names[1]
```

```
[15]: 'ALL CAP GROWTH FUND'
```

```python
[16]: common_summaries[1]
```

```
[16]: 'Fund Summary\nInvestment Objective\nThe fund seeks long-term capital
      growth.\nFees and Expenses\nThe following table describes the fees and expenses
      you may pay if you buy and hold shares of the fund. You may qualify for sales
      charge discounts if you and your family invest, or agree to invest in the
      future, at least $50,000 in American Century Investments funds. More information
      about these and other discounts is available from your financial professional
```

and in Calculation of Sales Charges on page 12 of the fund's prospectus, Appendix A of the fund's prospectus and Sales Charges in Appendix B of the statement of additional information.\nShareholder Fees (fees paid directly from your investment)\n \n \nInvestor\nI1\nA\nC\nR\nMaximum Sales Charge (Load) Imposed on\nPurchases (as a percentage of offering price)\nNone\nNone\n5.75%\nNone\nNone\nMaximum Deferred Sales Charge (Load)\n(as a percentage of the lower of the original\noffering price or redemption proceeds when redeemed\nwithin one year of purchase)\nNone\nNone\nNone[2]\n1.00%\nNone\nMaximum Annual Account Maintenance Fee\n(waived for unmatured Giftrust accounts or if\neligible investments total at least $10,000)\n$25\nNone\nNone\nNone\nNone\nAnnual Fund Operating Expenses (expenses that you pay each year as a percentage of the value of your investment)\n \nInvestor\nI\nA\nC\nR\nManagement Fee\n1.00%\n0.80%\n1.00%\n1.00%\n1.00%\nDistribution and Service (12b-1) Fees\nNone\nNone\n0.25%\n1.00%\n0.50%\nOther Expenses\n0.01%\n0.01%\n0.01%\n0.01%\n0.01%\nTotal Annual Fund Operating Expenses\n1.01%\n0.81%\n1.26%\n2.01%\n1.51%\n1 \nPrior to April 10, 2017, this class was referred to as the Institutional Class.\n2 \nPurchases of $1 million or more may be subject to a contingent deferred sales charge of 1.00% if the shares are redeemed within one year of the date of the purchase.\nExample\nThe example below is intended to help you compare the costs of investing in the fund with the costs of investing in other mutual funds. The example assumes that you invest $10,000 in the fund for the time periods indicated and then redeem all of your shares at the end of those periods, that you earn a 5% return each year, and that the fund's operating expenses remain the same. You may be required to pay brokerage commissions on your purchases of Investor or I Class shares of the fund, which are not reflected in the example. Although your actual costs may be higher or lower, based on these assumptions your costs would be:\n \n\n1 year\n3 years\n5 years\n10 years\nInvestor Class\n$103\n$322\n$559\n$1,236\nI Class\n$83\n$259\n$450\n$1,002\nA Class\n$696\n$952\n$1,228\n$2,010\nC Class\n$204\n$632\n$1,084\n$2,334\nR Class\n$154\n$478\n$824\n$1,800\nPortfolio Turnover\nThe fund pays transaction costs, such as commissions, when it buys and sells securities (or "turns over" its portfolio). A higher portfolio turnover rate may indicate higher transaction costs and may result in higher taxes when fund shares are held in a taxable account. These costs, which are not reflected in annual fund operating expenses or in the example, affect the fund's performance. During the most recent fiscal year, the fund's portfolio turnover rate was 44% of the average value of its portfolio.\n\n2#\n\n\n\n\nPrincipal Investment Strategies \nThe portfolio managers look for stocks of companies of all sizes they believe will increase in value over time, using an investment strategy developed by the fund's investment advisor. In implementing this strategy, the portfolio managers make their investment decisions based primarily on their analysis of individual companies, rather than on broad economic forecasts. The portfolio managers use a variety of analytical research tools and techniques to identify companies that meet their investment criteria. Under normal market conditions, the fund's portfolio will primarily consist of securities of companies demonstrating business improvement. Analytical

indicators helping to identify signs of business improvement could include accelerating earnings or revenue growth rates, increasing cash flows, or other indications of the relative strength of a company's business. These techniques help the portfolio managers buy or hold the stocks of companies they believe have favorable growth prospects and sell the stocks of companies whose characteristics no longer meet their criteria.\nAlthough the portfolio managers intend to invest the fund's assets primarily in U.S. securities, the fund may invest in securities of foreign companies when these securities meet the portfolio managers' standards of selection.\nThe fund may write covered calls on a portion of the fund's holdings in common stock when the portfolio managers believe call premiums are attractive relative to the price of the underlying security.\nPrincipal Risks \n•\nGrowth Stocks – Investments in growth stocks may be more volatile than other stocks and the overall stock market. These stocks are typically priced higher than other stocks because of their growth potential, which may or may not be realized.\n•\nSmall and Mid Cap Stocks – The fund invests in mid-sized and smaller companies, which may be more volatile and subject to greater risk than larger companies. Smaller companies may have limited financial resources, product lines and markets, and their securities may trade less frequently and in more limited volumes than the securities of larger companies, which could lead to higher transaction costs.\n•\nStyle Risk – If at any time the market is not favoring the fund's growth investment style, the fund's gains may not be as big as, or its losses may be bigger than, those of other equity funds using different investment styles.\n•\nForeign Securities – The fund may invest in foreign securities, which can be riskier than investing in U.S. securities. Securities of foreign issuers may be less liquid, more volatile and harder to value than U.S. securities.\n•\nCovered Call Risk – Writing covered calls may limit the fund's ability to participate in price increases of the underlying securities and could lower the fund's return.\n•\nPrice Volatility – The value of the fund's shares may fluctuate significantly in the short term.\n•\nMarket Risk – The value of the fund's shares will go up and down based on the performance of the companies whose securities it owns and other factors generally affecting the securities market.\n•\nPrincipal Loss – At any given time your shares may be worth less than the price you paid for them. In other words, it is possible to lose money by investing in the fund.\nAn investment in the fund is not a bank deposit, and it is not insured or guaranteed by the Federal Deposit Insurance Corporation (FDIC) or any other government agency.\nFund Performance \nThe following bar chart and table provide some indication of the risks of investing in the fund. The bar chart shows changes in the fund's performance from year to year for Investor Class shares. The table shows how the fund's average annual returns for the periods shown compared with those of a broad measure of market performance. The fund's past performance (before and after taxes) is not necessarily an indication of how the fund will perform in the future. For current performance information, please visit americancentury.com. \nSales charges and account fees, if applicable, are not reflected in the bar chart. If those charges were included, returns would be less than those shown.\n\n3#\n\n\n\n\nCalendar Year Total Returns \nchart-4e88d1b1ec1288e862aa01.jpg\nHighest Performance Quarter
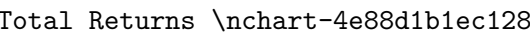
(1Q 2012): 18.71% Lowest Performance Quarter (4Q 2008): -26.56%\nAverage Annual Total Returns \nFor the calendar year ended December 31, 2017\n1 year\n5 years\n10 years\nSince Inception\nInception Date\nInvestor Class Return Before Taxes\n24.87%\n14.13%\n6.76%\n-\n11/25/1983\nReturn After Taxes on Distributions\n22.57%\n11.48%\n5.36%\n-\n11/25/1983\nReturn After Taxes on Distributions and Sale of Fund Shares\n15.98%\n10.85%\n5.22%\n-\n11/25/1983\nI Class Return Before Taxes\n25.14%\n14.36%\n-\n15.32%\n09/30/2011\nA Class1 Return Before Taxes\n17.42%\n12.51%\n5.87%\n-\n09/30/2011\nC Class1 Return Before Taxes\n23.63%\n12.99%\n5.70%\n-\n09/30/2011\nR Class1 Return Before Taxes\n24.24%\n13.57%\n6.23%\n-\n09/30/2011\nRussell 3000® Growth Index\n (reflects no deduction for fees, expenses or taxes)\n29.59%\n17.15%\n9.93%\n-\n-\n \n1 \nHistorical performance for A, C and R Classes prior to their inception is based on the performance of Investor Class shares. A, C and R Class performance has been adjusted to reflect differences in sales charges, if applicable, and expenses between classes.\nThe after-tax returns are shown only for Investor Class shares. After-tax returns for other share classes will vary. After-tax returns are calculated using the historical highest individual federal marginal income tax rates and do not reflect the impact of state and local taxes. Actual after-tax returns depend on an investor's tax situation and may differ from those shown. After-tax returns are not relevant to investors who hold their fund shares through tax-deferred arrangements, such as 401(k) plans or IRAs.\nPortfolio Management \nInvestment Advisor \nAmerican Century Investment Management, Inc. \nPortfolio Managers \nGregory J. Woodhams , CFA, Co-CIO, Global Growth Equity, Senior Vice President and Senior Portfolio Manager, has been a member of the team that manages the fund since 2018.\nJoseph Reiland , CFA, Vice President and Portfolio Manager, has been a member of the team that manages the fund since 2018.\n\n4#\n\n\n\n\n\nPurchase and Sale of Fund Shares \nYou may purchase or redeem shares of the fund on any business day through our website at americancentury.com, in person (at one of our Investor Centers), by mail (American Century Investments, P.O. Box 419200, Kansas City, MO 64141-6200), by telephone at 1-800-345-2021 (Investor Services Representative) or 1-800-345-3533 (Business, Not-For-Profit and Employer-Sponsored Retirement Plans), or through a financial intermediary. Shares may be purchased and redemption proceeds received by electronic bank transfer, by check or by wire. \nUnless otherwise specified below, the minimum initial investment amount to open an account is $2,500 ($2,000 for Coverdell Education Savings Accounts). Investors opening accounts through financial intermediaries may open an account with $250 for Investor, A, C and R Classes, but the financial intermediaries may require their clients to meet different investment minimums. The minimum may be waived for broker-dealer sponsored wrap program accounts, fee based accounts, and accounts through bank/trust and wealth management advisory organizations. \nThe minimum initial investment amount for I Class is generally $5 million ($3 million for endowments and foundations), but the minimum may be waived if you have an aggregate investment in the American Century family of funds of $10 million or more ($5 million for endowments and foundations). This includes accounts held directly with American Century and those held through a financial intermediary. \nFor Investor, A, C and R Classes, there is no minimum

initial investment amount for certain employer-sponsored retirement plans,
however, financial intermediaries or plan recordkeepers may require plans to
meet different minimums. For purposes of fund minimums, employer-sponsored
retirement plans do not include SEP IRAs, SIMPLE IRAs or SARSEPs. Employer-
sponsored retirement plans are not eligible to invest in the I Class.\nThere is
a $50 minimum for subsequent purchases, except there is no subsequent purchase
minimum for financial intermediaries or employer-sponsored retirement plans.
\nPurchase and Sale of Fund Shares Through a Giftrust® \nPrior to June 21,
2017, All Cap Growth shares could be purchased to establish a Giftrust, which
was a product created by American Century as a way to invest in the fund for
someone other than yourself through a one-time gift held in an irrevocable
trust.\nAs of June 21, 2017, no new Giftrusts may be opened. However, additional
shares may be purchased for Giftrusts established prior to August 1, 2002, by
check or by wire.\nOnce a Giftrust matures, the shares are transferred out of
the Giftrust and into a new account registered in the beneficiary's individual
name. The beneficiary may purchase additional shares or redeem shares in the new
account as indicated above in the Purchase and Sale of Fund Shares section.\nTax
Information \nFund distributions are generally taxable as ordinary income or
capital gains, unless you are investing through a tax-deferred account such as a
401(k) or individual retirement account (in which case you may be taxed upon
withdrawal of your investment from such account). \nPayments to Broker-Dealers
and Other Financial Intermediaries \nIf you purchase the fund through a broker-
dealer or other financial intermediary (such as a bank, insurance company, plan
sponsor or financial professional), the fund and its related companies may pay
the intermediary for the sale of fund shares and related services. These
payments may create a conflict of interest by influencing the broker-dealer or
other intermediary and your salesperson to recommend the fund over another
investment. Ask your salesperson or visit your financial intermediary's website
for more information.'

0.2 split to train, validation, test

```
[17]: X_train, X_val_test, y_train, y_val_test = train_test_split(
          common_summaries, common_fund_names, test_size=0.4, random_state=815)
```

```
[18]: #split to test and validation set
      X_val, X_test, y_val, y_test = train_test_split(
          X_val_test, y_val_test, test_size=0.5, random_state=815)
```

```
[19]: len(X_train), len(X_val), len(X_test)
```

```
[19]: (259, 86, 87)
```

```
[20]: y_train
```

```
[20]: ['MFS Maryland Municipal Bond Fund',
       'Janus Henderson Enterprise Portfolio',
       'Ivy Municipal Bond Fund',
       'Fidelity Advisor New Insights Fund',
       'CAPITAL VALUE FUND',
       'Franklin FTSE Mexico ETF',
       'Dreyfus Global Equity Income Fund',
       'Fidelity Emerging Markets Discovery Fund',
       'Short-Term Bond Fund',
       'Capital Income Builder',
       'Janus Henderson Balanced Fund',
       'Federated MDT Large Cap Growth Fund',
       'Ultra-Short Bond Fund',
       'Ivy High Income Fund',
       'Treasury Money Market Trust',
       'Fidelity Advisor Energy Fund',
       'MainStay VP International Equity Portfolio',
       'Columbia Variable Portfolio - Emerging Markets Fund',
       'American Funds College 2018 Fund',
       'PIMCO California Intermediate Municipal Bond Fund',
       'Columbia Variable Portfolio - Select Large-Cap Value Fund',
       'MainStay VP Balanced Portfolio',
       'Columbia Flexible Capital Income Fund',
       'Janus Henderson Global Unconstrained Bond Portfolio',
       'PIMCO REALPATH 2055',
       'NORTHERN U.S. GOVERNMENT SELECT MONEY MARKET FUND',
       'Eaton Vance Large-Cap Value Fund',
       'Ivy Value Fund',
       'T. Rowe Price Africa & Middle East Fund',
       'Ivy Global Equity Income Fund',
       'Fidelity Advisor Utilities Fund',
       'American Funds Tax-Advantaged Growth and Income Portfolio',
       'Fidelity Emerging Asia Fund',
       'Vanguard Intermediate-Term Corporate Bond Index Fund',
       'T. Rowe Price Emerging Markets Value Stock Fund',
       'Fidelity Series Short-Term Credit Fund',
       'Vanguard Energy Fund',
       'Growth Fund',
       'Columbia Variable Portfolio - Global Bond Fund',
       'American Century STOXX U.S. Quality Value ETF',
       'Janus Henderson Global Technology Fund',
       'Blue Chip Income and Growth Fund',
       'Columbia Variable Portfolio - Balanced Fund',
       'Janus Henderson International Opportunities Fund',
       'Fidelity Contrafund',
       'American Funds 2055 Target Date Retirement Fund',
       'MainStay VP MFS Utilities Portfolio',
```

'Fidelity Advisor Semiconductors Fund',
'Franklin Payout 2023 Fund',
'Ivy Securian Core Bond Fund',
'Columbia Small Cap Value Fund II',
'American Funds 2020 Target Date Retirement Fund',
'Franklin FTSE Hong Kong ETF',
'Columbia Select Large Cap Equity Fund',
'Vanguard Long-Term Corporate Bond Index Fund',
'Janus Henderson Flexible Bond Portfolio',
'Variable Portfolio - Oppenheimer International Growth Fund',
'Fidelity Advisor Freedom 2015 Fund',
'Franklin LibertyQ U.S. Small Cap Equity ETF',
'Dreyfus International Bond Fund',
'Franklin FTSE Italy ETF',
'American Century Diversified Corporate Bond ETF',
'MainStay VP Common Stock Portfolio',
'Fidelity Flex U.S. Bond Index Fund',
'T. Rowe Price Emerging Markets Stock Fund',
'Franklin FTSE Asia ex Japan ETF',
'AllianzGI Health Sciences Fund',
'PIMCO REALPATH 2030 Fund',
'Schwab New York Municipal Money Fund',
'Ivy Limited-Term Bond Fund',
'Franklin FTSE Australia ETF',
'Columbia Mid Cap Index Fund',
'SMALL CAP GROWTH FUND',
'Janus Henderson Global Technology Portfolio',
'T. Rowe Price Emerging Europe Fund',
'Portfolio Series - American Funds Managed Risk Global Allocation Portfolio',
'MainStay VP Income Builder Portfolio',
'American Funds College 2033 Fund',
'American Funds 2010 Target Date Retirement Fund',
'Fidelity Advisor Technology Fund',
'Janus Henderson U.S. Growth Opportunities Fund',
'JPMorgan Tax Aware Equity Fund',
'Columbia Variable Portfolio - Large Cap Index Fund',
'Janus Henderson Mid Cap Value Portfolio',
'Fidelity Emerging Europe, Middle East, Africa (EMEA) Fund',
'ULTRA FUND',
'JPMorgan Europe Dynamic Fund',
'Franklin FTSE Russia ETF',
'PIMCO REALPATH 2045 Fund',
'Vanguard Real Estate II Index Fund',
'American Funds Retirement Income Portfolio - Enhanced',
'BALANCED FUND',
'American Century Value ETF',
'Ivy Small Cap Growth Fund',

'Cornerstone Moderately Conservative Fund',
'Managed Risk Growth Fund',
'Eaton Vance Growth Fund',
'Portfolio Series - American Funds Growth and Income Portfolio',
'Fidelity Managed Retirement Income Fund',
'Ivy Small Cap Core Fund',
'Vanguard Russell 1000 Growth Index Fund',
'Franklin FTSE Europe Hedged ETF',
'Vanguard Intermediate-Term Investment-Grade Fund',
'Columbia Seligman Communications and Information Fund',
'Columbia Variable Portfolio - High Yield Bond Fund',
'Dreyfus Global Emerging Markets Fund',
'American Funds College 2021 Fund',
'MFS Global Growth Fund',
'Vanguard Short-Term Investment-Grade Fund',
'Portfolio Series - American Funds Managed Risk Growth and Income Portfolio',
'American Funds College 2024 Fund',
'American Funds College 2036 Fund',
'Ivy Managed International Opportunities Fund',
'Janus Henderson European Focus Fund',
'Columbia Variable Portfolio - Government Money Market Fund',
'NORTHERN MUNICIPAL MONEY MARKET FUND',
'Fidelity Contrafund K6',
'JPMorgan Equity Focus Fund',
'AllianzGI Focused Growth Fund',
'ADAPTIVE EQUITY FUND',
'Fidelity Emerging Markets Fund',
'USAA MSCI USA Small Cap Value Momentum Blend Index ETF',
'Janus Henderson Emerging Markets Fund',
'Franklin FTSE United Kingdom ETF',
'American Funds Moderate Growth and Income Portfolio',
'Fidelity GNMA Fund',
'Franklin Payout 2021 Fund',
'Franklin Liberty Intermediate Municipal Opportunities ETF',
'American Century Growth ETF',
'DELAWARE GLOBAL VALUE FUND',
'Fidelity Advisor Freedom 2020 Fund',
'JPMorgan New York Tax Free Bond Fund',
'Janus Henderson Overseas Portfolio',
'MainStay VP Mid Cap Core Portfolio',
'Western Asset Intermediate Maturity California Municipals Fund',
'Vanguard Ultra-Short-Term Bond Fund',
'Fidelity ZERO Extended Market Index Fund',
'Franklin FTSE China ETF',
'MFS South Carolina Municipal Bond Fund',
'CALIFORNIA HIGH-YIELD MUNICIPAL FUND',
'Federated MDT Balanced Fund',

'Fidelity Flex Short-Term Bond Fund',
'Small Cap Stock Fund',
'PIMCO REALPATH 2025 Fund',
'Eaton Vance New Jersey Municipal Income Fund',
'Eaton Vance TABS Short-Term Municipal Bond Fund',
'Ivy Asset Strategy Fund',
'PIMCO California Municipal Bond Fund',
'MainStay VP Emerging Markets Equity Portfolio',
'MFS Mississippi Municipal Bond Fund',
'Vanguard Real Estate Index Fund',
'Columbia Large Cap Value Fund',
'SUSTAINABLE EQUITY FUND',
'MainStay VP Eagle Small Growth Portfolio',
'Fidelity Series Corporate Bond Fund',
'Vanguard Short-Term Federal Fund',
'MainStay VP Epoch U.S. Small Cap Portfolio',
'Franklin Liberty U.S. Low Volatility ETF',
'Eaton Vance Global Small-Cap Fund',
'Fidelity Advisor Health Care Fund',
'Fidelity Advisor Freedom 2030 Fund',
'Franklin LibertyQ Global Dividend ETF',
'Vanguard Long-Term Treasury Index Fund',
'Janus Henderson Contrarian Fund',
'Ivy Micro Cap Growth Fund',
'Vanguard Russell 2000 Index Fund',
'Variable Portfolio - MFS Value Fund',
'AllianzGI Small-Cap Fund',
'MFS Utilities Fund',
'Ivy Tax-Managed Equity Fund',
'Total Return Strategy Fund',
'AllianzGI Income & Growth Fund',
'American Funds 2045 Target Date Retirement Fund',
'Variable Portfolio - Morgan Stanley Advantage Fund',
'USAA MSCI Emerging Markets Value Momentum Blend Index ETF',
'PIMCO REALPATH 2035 Fund',
'Janus Henderson International Value Fund',
'Janus Henderson International Small Cap Fund',
'Franklin FTSE South Korea ETF',
'MFS Global Equity Fund',
'DELAWARE EMERGING MARKETS FUND',
'International Growth and Income Fund',
'Columbia Large Cap Index Fund',
'Eaton Vance TABS 10-to-20 Year Laddered Municipal Bond Fund',
'PIMCO REALPATH 2050 Fund',
'Columbia Variable Portfolio - Mid Cap Growth Fund',
'Fidelity Advisor Freedom 2060 Fund',
'MFS New York Municipal Bond Fund',

'Vanguard Intermediate-Term Treasury Fund',
'Columbia Mid Cap Value Fund',
'Fidelity Managed Retirement 2025 Fund',
'Janus Henderson Global Research Fund',
'American Century Diversified Municipal Bond ETF',
'Columbia Variable Portfolio - Large Cap Growth Fund',
'American Funds 2030 Target Date Retirement Fund',
'Growth and Tax Strategy Fund',
'Janus Henderson Global Bond Portfolio',
'American Funds Tax-Exempt Preservation Portfolio',
'NT GROWTH FUND',
'Ivy Municipal High Income Fund',
'Delaware International Small Cap Fund',
'Vanguard Health Care Fund',
'Federated MDT Small Cap Core Fund',
'AllianzGI NFJ International Value Fund',
'Franklin FTSE Brazil ETF',
'Columbia Select International Equity Fund',
'Ivy Emerging Markets Equity Fund',
'MFS Municipal Income Fund',
'Vanguard Short-Term Treasury Index Fund',
'Columbia Variable Portfolio - Intermediate Bond Fund',
'Fidelity Advisor Freedom 2055 Fund',
'Fidelity Advisor Industrials Fund',
'Fidelity Simplicity RMD 2015 Fund',
'Managed Risk International Fund',
'Eaton Vance Municipal Opportunities Fund',
'Variable Portfolio - MFS Blended Research Core Equity Fund',
'JPMorgan Global Allocation Fund',
'PIMCO Gurtin California Municipal Opportunistic Value Fund',
'AllianzGI NFJ Mid-Cap Value Fund',
'Asset Allocation Fund',
'MFS Tennessee Municipal Bond Fund',
'CALIFORNIA INTERMEDIATE-TERM TAX-FREE BOND FUND',
'Fidelity Advisor Communications Equipment Fund',
'Government Securities Fund',
'Eaton Vance TABS 1-to-10 Year Laddered Municipal Bond Fund',
'Fidelity Simplicity RMD 2020 Fund',
'American Funds Retirement Income Portfolio - Moderate',
'Variable Portfolio - Wells Fargo Short Duration Government Fund',
'Ivy Mid Cap Growth Fund',
'Fidelity Municipal Income 2021 Fund',
'Variable Portfolio - TCW Core Plus Bond Fund',
'Columbia Select Global Growth Fund',
'Schwab California Municipal Money Fund',
'Fidelity Advisor Series Equity Value Fund',
'USAA MSCI USA Value Momentum Blend Index ETF',

```
'Eaton Vance Pennsylvania Municipal Income Fund',
'HERITAGE FUND',
'DELAWARE INTERNATIONAL VALUE EQUITY FUND',
'MainStay VP T. Rowe Price Equity Income Portfolio',
'Global Growth Fund',
'Janus Henderson Growth and Income Fund',
'Janus Henderson Forty Fund',
'Columbia Variable Portfolio - Select Smaller-Cap Value Fund',
'Janus Henderson Global Equity Income Fund',
'MainStay VP Small Cap Core Portfolio',
'Vanguard Intermediate-Term Treasury Index Fund',
'Eaton Vance Small-Cap Fund',
'Western Asset Massachusetts Municipals Fund',
'Eaton Vance Special Equities Fund',
'MFS North Carolina Municipal Bond Fund',
'Janus Henderson Global Research Portfolio',
'Franklin FTSE Taiwan ETF',
'Fidelity Advisor Freedom 2040 Fund',
'Ivy Global Income Allocation Fund',
'Intermediate-Term Bond Fund',
'PIMCO California Short Duration Municipal Income Fund',
'Fidelity Simplicity RMD 2005 Fund',
'Variable Portfolio - Victory Sycamore Established Value Fund',
'Franklin Payout 2019 Fund']
```

0.3 Following the NLP application in class, use the skip-gram model to build a word embedding dictionary from the mutual fund prospectus in the training set

0.3.1 Tokenize the text

```python
[21]: def lemma_tokenizer(text):
          lemmatizer=WordNetLemmatizer()
          return [lemmatizer.lemmatize(token) for token in word_tokenize(text.
      ↪replace("'"," "))]
```

```python
[22]: stop_words = set(stopwords.words('english')+list(string.
      ↪punctuation)+['``',"'''"]+["]","[","*"]+['doe', 'ha', 'wa'])
```

```python
[23]: # clean and tokenize the text
      def tokenizer(txt):
          txt = txt.replace('\n', ' ').replace('\t', ' ').lower()
          word_tokens = word_tokenize(txt)
          filtered_sentence = [w for w in word_tokens if not w in stop_words]
          filtered_sentence = [w for w in filtered_sentence if re.sub("[^A-Za-z
      ↪]+",'',w) != '']
          return filtered_sentence
```

```
[ ]: #extract X_train as a list
     #X_train_list = X_train.tolist()
```

```
[ ]: #import nltk
     #nltk.download('punkt')
```

```
[ ]: #nltk.download('punkt_tab')
```

```
[24]: text_words = np.concatenate([tokenizer(summary) for summary in X_train])
```

```
[25]: text_words, text_words.size
```

```
[25]: (array(['mfs', 'maryland', 'municipal', …, "'s", 'website',
              'information'], dtype='<U44'),
       464939)
```

0.3.2 Process Skip-Gram model Input

```
[26]: # Training Parameters
      batch_size = 128 # The model will be trained batch per batch and one batch␣
       ↪contains 128 rows
      num_epochs = 2 # The model will go through all the data twice
```

```
[27]: # Word2Vec Parameters
      embedding_size = 50 # Dimension of the embedding vector
      max_vocabulary_size = 5000 # Total number of different words in the vocabulary
      min_occurrence = 10 # Remove all words that does not appears at least n times
      skip_window = 3 # How many words to consider left and right
      num_skips = 4 # How many times to reuse an input to generate a label
```

```
[28]: # Build the dictionary and replace rare words with UNK token
      count = [('UNK', -1)]
      # Retrieve the most common words
      count.extend(collections.Counter(text_words).most_common(max_vocabulary_size -␣
       ↪1))
      # Remove samples with less than 'min_occurrence' occurrences
      for i in range(len(count) - 1, -1, -1):
          if count[i][1] < min_occurrence:
              count.pop(i)
          else:
              # The collection is ordered, so stop when 'min_occurrence' is reached
              break
```

```
[29]: word2id = dict() #same as {}
      for i, (word, _)in enumerate(count):
          word2id[word] = i
```

```
#word2id
```

```python
id2word = dict(zip(word2id.values(), word2id.keys()))
vocab_size = len(id2word)
```

```
#id2word
```

```python
print ('size of the vocabulary : '+str(vocab_size))
```

size of the vocabulary : 2699

```python
len(word2id) #the words in'id2word and word2id is same
```

2699

```python
# create data
data = list() #[]
unk_count = 0
for word in text_words:
    # Retrieve a word id, or assign it index 0 ('UNK') if not in dictionary
    index = word2id.get(word, 0) #if vwe cannot find the word, it will return 0
    instead of having an error
    if index == 0:  #which means we could not find the word in our dictionary
        unk_count += 1
    data.append(index)
count[0] = ('UNK', unk_count) #update the number of 'UNK'
```

```python
count[0]
```

('UNK', 9345)

```python
print(data[:20]) #377 is janus
```

[400, 0, 117, 138, 1, 292, 1347, 30, 4, 128, 1, 4, 128, 362, 39, 28, 1240, 25, 554, 78]

```python
# build OneHot vector from index
def to_one_hot(data_point_index, vocab_size):
    temp = np.zeros(vocab_size)
    temp[data_point_index] = 1
    return temp
```

```python
# Generate training batch for the skip-gram model
def batch_generator(batch_size, num_skips, skip_window, vocab_size):
    data_index = 0
    while True :
        assert batch_size % num_skips == 0
```

```python
        assert num_skips <= 2 * skip_window
        # batch is filled with 128 inputs
        batch = np.ndarray(shape=(batch_size), dtype=np.int32)
        # labels is filled with 128 outputs
        labels = np.ndarray(shape=(batch_size), dtype=np.int32)
        span = 2 * skip_window + 1
        # buffer keep track of the visited indexes visited
        buffer = collections.deque(maxlen=span)
        if data_index + span > len(data):
            data_index = 0
            # We stop the loop when we went through all the corpus
            break
        buffer.extend(data[data_index:data_index + span])
        data_index += span
        for i in range(batch_size // num_skips):
            # Take the context current word
            context_words = [w for w in range(span) if w != skip_window]
            # Randomly select num_skips words in the context
            words_to_use = random.sample(context_words, num_skips)
            for j, context_word in enumerate(words_to_use):
                # Creates one raw data
                batch[i * num_skips + j] = buffer[skip_window]
                labels[i * num_skips + j] = buffer[context_word]
            if data_index == len(data):
                buffer.extend(data[0:span])
                data_index = span
            else:
                buffer.append(data[data_index])
                data_index += 1
        # Backtrack a little bit to avoid skipping words in the end of a batch
        data_index = (data_index + len(data) - span) % len(data)

        # translate word index to on-hot representation
        batch_one_hot = np.array([to_one_hot(b, vocab_size) for b in batch])
        labels_one_hot = np.array([to_one_hot(l, vocab_size) for l in labels])

        # output one batch
        yield batch_one_hot, labels_one_hot
```

```
[ ]:
```

0.3.3 Train the skip-gram model

```python
[38]: # Create en compile the Autoencoder
      def creat_word2vec_model():
          input_word = Input(shape=(vocab_size,))
```

```python
    encoded = Dense(embedding_size, activation='linear')(input_word)
    decoded = Dense(vocab_size, activation='softmax')(encoded)

    # The autoencoder is the whole model with hidden layer contected to the
 ↪output layer.
    autoencoder = Model(input_word, decoded)
    # The encoder is just the input layer connected to the hidden layer. One
 ↪the Autoencoder will be trained we will use
    # the encoder to create our word vectors
    encoder = Model(input_word, encoded)

    autoencoder.compile(optimizer='adam', loss='binary_crossentropy')
    return encoder, autoencoder
```

```python
[39]: # We create the model
      encoder, autoencoder = creat_word2vec_model()
```

```python
[40]: autoencoder.summary()
```

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 2699)	0
dense (Dense)	(None, 50)	135,000
dense_1 (Dense)	(None, 2699)	137,649

Total params: 272,649 (1.04 MB)

Trainable params: 272,649 (1.04 MB)

Non-trainable params: 0 (0.00 B)

```python
[41]: # Finally we can train the model by feeding it with our batch generator !
      autoencoder.fit(batch_generator(batch_size, num_skips, skip_window,
       ↪vocab_size), steps_per_epoch=ceil(len(data) / batch_size), epochs=num_epochs)
```

```
Epoch 1/2
3633/3633                25s 7ms/step -
loss: 0.0933
```

```
Epoch 2/2
3633/3633              27s 7ms/step -
loss: 0.0028
```

[41]: `<keras.src.callbacks.history.History at 0x7a92d00f30d0>`

[42]:
```python
# Create the Vectorizer function (prediciton of the encoder)
def vecotrize(word):
    word_one_hot = to_one_hot(word2id[word], vocab_size)
    return encoder.predict(np.array([word_one_hot]))[0]
```

[43]:
```python
# Create the word2vec dictionary
word2vec = {w : vecotrize(w) for w in word2id.keys()}

# This dictionary gives for all words it's vectorial representation.
```

```
1/1                    0s 157ms/step
1/1                    0s 29ms/step
1/1                    0s 43ms/step
1/1                    0s 39ms/step
1/1                    0s 40ms/step
1/1                    0s 40ms/step
1/1                    0s 42ms/step
1/1                    0s 47ms/step
1/1                    0s 43ms/step
1/1                    0s 38ms/step
1/1                    0s 42ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 41ms/step
1/1                    0s 55ms/step
1/1                    0s 39ms/step
1/1                    0s 39ms/step
1/1                    0s 42ms/step
1/1                    0s 44ms/step
1/1                    0s 44ms/step
1/1                    0s 46ms/step
1/1                    0s 41ms/step
1/1                    0s 47ms/step
1/1                    0s 42ms/step
1/1                    0s 40ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
```

```
1/1                0s 29ms/step
1/1                0s 27ms/step
1/1                0s 27ms/step
1/1                0s 27ms/step
1/1                0s 29ms/step
1/1                0s 28ms/step
1/1                0s 27ms/step
1/1                0s 28ms/step
1/1                0s 28ms/step
1/1                0s 28ms/step
1/1                0s 28ms/step
1/1                0s 31ms/step
1/1                0s 27ms/step
1/1                0s 28ms/step
1/1                0s 27ms/step
1/1                0s 29ms/step
1/1                0s 29ms/step
1/1                0s 29ms/step
1/1                0s 27ms/step
1/1                0s 27ms/step
1/1                0s 27ms/step
1/1                0s 28ms/step
1/1                0s 32ms/step
1/1                0s 28ms/step
1/1                0s 28ms/step
1/1                0s 29ms/step
1/1                0s 27ms/step
1/1                0s 27ms/step
1/1                0s 27ms/step
1/1                0s 29ms/step
1/1                0s 28ms/step
1/1                0s 28ms/step
1/1                0s 29ms/step
1/1                0s 30ms/step
1/1                0s 28ms/step
1/1                0s 29ms/step
1/1                0s 30ms/step
1/1                0s 32ms/step
1/1                0s 30ms/step
1/1                0s 28ms/step
1/1                0s 30ms/step
1/1                0s 31ms/step
1/1                0s 27ms/step
1/1                0s 37ms/step
1/1                0s 29ms/step
1/1                0s 32ms/step
1/1                0s 30ms/step
1/1                0s 29ms/step
```

```
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 27ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 38ms/step
1/1                    0s 28ms/step
1/1                    0s 31ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 27ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 35ms/step
1/1                    0s 28ms/step
1/1                    0s 31ms/step
1/1                    0s 35ms/step
1/1                    0s 28ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
```

```
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 34ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 41ms/step
1/1                    0s 29ms/step
1/1                    0s 27ms/step
1/1                    0s 27ms/step
1/1                    0s 27ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 44ms/step
1/1                    0s 28ms/step
1/1                    0s 27ms/step
1/1                    0s 27ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
```

```
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 28ms/step
1/1                    0s 37ms/step
1/1                    0s 47ms/step
1/1                    0s 47ms/step
1/1                    0s 48ms/step
1/1                    0s 45ms/step
1/1                    0s 48ms/step
1/1                    0s 44ms/step
1/1                    0s 44ms/step
1/1                    0s 42ms/step
1/1                    0s 40ms/step
1/1                    0s 49ms/step
1/1                    0s 51ms/step
1/1                    0s 45ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 43ms/step
1/1                    0s 42ms/step
1/1                    0s 42ms/step
1/1                    0s 48ms/step
1/1                    0s 48ms/step
1/1                    0s 53ms/step
1/1                    0s 42ms/step
1/1                    0s 40ms/step
1/1                    0s 40ms/step
1/1                    0s 50ms/step
1/1                    0s 31ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 41ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 36ms/step
1/1                    0s 32ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
```

```
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 37ms/step
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 28ms/step
1/1                    0s 28ms/step
1/1                    0s 32ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 28ms/step
1/1                    0s 32ms/step
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 38ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 32ms/step
1/1                    0s 29ms/step
1/1                    0s 27ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 42ms/step
1/1                    0s 32ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 28ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
```

```
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 35ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 28ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 44ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
```

```
1/1          0s 30ms/step
1/1          0s 30ms/step
1/1          0s 35ms/step
1/1          0s 32ms/step
1/1          0s 29ms/step
1/1          0s 29ms/step
1/1          0s 30ms/step
1/1          0s 29ms/step
1/1          0s 38ms/step
1/1          0s 33ms/step
1/1          0s 34ms/step
1/1          0s 30ms/step
1/1          0s 35ms/step
1/1          0s 29ms/step
1/1          0s 35ms/step
1/1          0s 32ms/step
1/1          0s 30ms/step
1/1          0s 29ms/step
1/1          0s 31ms/step
1/1          0s 29ms/step
1/1          0s 30ms/step
1/1          0s 34ms/step
1/1          0s 40ms/step
1/1          0s 32ms/step
1/1          0s 30ms/step
1/1          0s 45ms/step
1/1          0s 46ms/step
1/1          0s 46ms/step
1/1          0s 43ms/step
1/1          0s 47ms/step
1/1          0s 49ms/step
1/1          0s 39ms/step
1/1          0s 65ms/step
1/1          0s 44ms/step
1/1          0s 40ms/step
1/1          0s 40ms/step
1/1          0s 39ms/step
1/1          0s 44ms/step
1/1          0s 46ms/step
1/1          0s 41ms/step
1/1          0s 47ms/step
1/1          0s 45ms/step
1/1          0s 44ms/step
1/1          0s 39ms/step
1/1          0s 42ms/step
1/1          0s 49ms/step
1/1          0s 45ms/step
1/1          0s 50ms/step
```

```
1/1                     0s 42ms/step
1/1                     0s 50ms/step
1/1                     0s 49ms/step
1/1                     0s 49ms/step
1/1                     0s 32ms/step
1/1                     0s 30ms/step
1/1                     0s 31ms/step
1/1                     0s 32ms/step
1/1                     0s 30ms/step
1/1                     0s 32ms/step
1/1                     0s 35ms/step
1/1                     0s 35ms/step
1/1                     0s 30ms/step
1/1                     0s 30ms/step
1/1                     0s 32ms/step
1/1                     0s 43ms/step
1/1                     0s 31ms/step
1/1                     0s 32ms/step
1/1                     0s 32ms/step
1/1                     0s 30ms/step
1/1                     0s 34ms/step
1/1                     0s 37ms/step
1/1                     0s 31ms/step
1/1                     0s 32ms/step
1/1                     0s 33ms/step
1/1                     0s 33ms/step
1/1                     0s 31ms/step
1/1                     0s 34ms/step
1/1                     0s 31ms/step
1/1                     0s 31ms/step
1/1                     0s 29ms/step
1/1                     0s 29ms/step
1/1                     0s 36ms/step
1/1                     0s 34ms/step
1/1                     0s 32ms/step
1/1                     0s 34ms/step
1/1                     0s 31ms/step
1/1                     0s 31ms/step
1/1                     0s 31ms/step
1/1                     0s 30ms/step
1/1                     0s 30ms/step
1/1                     0s 29ms/step
1/1                     0s 32ms/step
1/1                     0s 35ms/step
1/1                     0s 29ms/step
1/1                     0s 30ms/step
1/1                     0s 31ms/step
1/1                     0s 34ms/step
```

```
1/1                   0s 31ms/step
1/1                   0s 31ms/step
1/1                   0s 30ms/step
1/1                   0s 32ms/step
1/1                   0s 33ms/step
1/1                   0s 30ms/step
1/1                   0s 30ms/step
1/1                   0s 30ms/step
1/1                   0s 29ms/step
1/1                   0s 34ms/step
1/1                   0s 33ms/step
1/1                   0s 31ms/step
1/1                   0s 31ms/step
1/1                   0s 30ms/step
1/1                   0s 30ms/step
1/1                   0s 32ms/step
1/1                   0s 31ms/step
1/1                   0s 31ms/step
1/1                   0s 29ms/step
1/1                   0s 30ms/step
1/1                   0s 32ms/step
1/1                   0s 36ms/step
1/1                   0s 33ms/step
1/1                   0s 47ms/step
1/1                   0s 34ms/step
1/1                   0s 34ms/step
1/1                   0s 31ms/step
1/1                   0s 31ms/step
1/1                   0s 30ms/step
1/1                   0s 34ms/step
1/1                   0s 31ms/step
1/1                   0s 30ms/step
1/1                   0s 31ms/step
1/1                   0s 36ms/step
1/1                   0s 31ms/step
1/1                   0s 30ms/step
1/1                   0s 33ms/step
1/1                   0s 32ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
1/1                   0s 30ms/step
1/1                   0s 32ms/step
1/1                   0s 30ms/step
1/1                   0s 31ms/step
1/1                   0s 32ms/step
1/1                   0s 32ms/step
1/1                   0s 33ms/step
1/1                   0s 33ms/step
```

```
1/1          0s 31ms/step
1/1          0s 33ms/step
1/1          0s 31ms/step
1/1          0s 29ms/step
1/1          0s 31ms/step
1/1          0s 32ms/step
1/1          0s 37ms/step
1/1          0s 30ms/step
1/1          0s 28ms/step
1/1          0s 32ms/step
1/1          0s 40ms/step
1/1          0s 30ms/step
1/1          0s 35ms/step
1/1          0s 39ms/step
1/1          0s 31ms/step
1/1          0s 33ms/step
1/1          0s 30ms/step
1/1          0s 29ms/step
1/1          0s 30ms/step
1/1          0s 33ms/step
1/1          0s 30ms/step
1/1          0s 32ms/step
1/1          0s 32ms/step
1/1          0s 31ms/step
1/1          0s 31ms/step
1/1          0s 36ms/step
1/1          0s 43ms/step
1/1          0s 29ms/step
1/1          0s 31ms/step
1/1          0s 32ms/step
1/1          0s 31ms/step
1/1          0s 33ms/step
1/1          0s 32ms/step
1/1          0s 34ms/step
1/1          0s 38ms/step
1/1          0s 33ms/step
1/1          0s 41ms/step
1/1          0s 46ms/step
1/1          0s 73ms/step
1/1          0s 39ms/step
1/1          0s 40ms/step
1/1          0s 41ms/step
1/1          0s 45ms/step
1/1          0s 42ms/step
1/1          0s 39ms/step
1/1          0s 46ms/step
1/1          0s 41ms/step
1/1          0s 51ms/step
```

```
1/1                    0s 44ms/step
1/1                    0s 45ms/step
1/1                    0s 49ms/step
1/1                    0s 42ms/step
1/1                    0s 39ms/step
1/1                    0s 39ms/step
1/1                    0s 41ms/step
1/1                    0s 45ms/step
1/1                    0s 45ms/step
1/1                    0s 46ms/step
1/1                    0s 48ms/step
1/1                    0s 44ms/step
1/1                    0s 53ms/step
1/1                    0s 46ms/step
1/1                    0s 50ms/step
1/1                    0s 33ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 30ms/step
1/1                    0s 34ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 32ms/step
1/1                    0s 29ms/step
1/1                    0s 31ms/step
1/1                    0s 34ms/step
1/1                    0s 30ms/step
1/1                    0s 36ms/step
1/1                    0s 31ms/step
1/1                    0s 33ms/step
1/1                    0s 36ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 33ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 29ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
```

```
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 30ms/step
1/1                    0s 34ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 38ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 32ms/step
1/1                    0s 37ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 29ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 30ms/step
1/1                    0s 31ms/step
1/1                    0s 38ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 32ms/step
1/1                    0s 29ms/step
1/1                    0s 30ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
```

```
1/1                     0s 32ms/step
1/1                     0s 31ms/step
1/1                     0s 32ms/step
1/1                     0s 36ms/step
1/1                     0s 30ms/step
1/1                     0s 30ms/step
1/1                     0s 31ms/step
1/1                     0s 31ms/step
1/1                     0s 29ms/step
1/1                     0s 32ms/step
1/1                     0s 35ms/step
1/1                     0s 36ms/step
1/1                     0s 32ms/step
1/1                     0s 33ms/step
1/1                     0s 32ms/step
1/1                     0s 32ms/step
1/1                     0s 31ms/step
1/1                     0s 31ms/step
1/1                     0s 34ms/step
1/1                     0s 31ms/step
1/1                     0s 33ms/step
1/1                     0s 32ms/step
1/1                     0s 34ms/step
1/1                     0s 34ms/step
1/1                     0s 34ms/step
1/1                     0s 32ms/step
1/1                     0s 32ms/step
1/1                     0s 41ms/step
1/1                     0s 31ms/step
1/1                     0s 32ms/step
1/1                     0s 31ms/step
1/1                     0s 35ms/step
1/1                     0s 30ms/step
1/1                     0s 34ms/step
1/1                     0s 33ms/step
1/1                     0s 35ms/step
1/1                     0s 33ms/step
1/1                     0s 36ms/step
1/1                     0s 44ms/step
1/1                     0s 34ms/step
1/1                     0s 35ms/step
1/1                     0s 35ms/step
1/1                     0s 45ms/step
1/1                     0s 50ms/step
1/1                     0s 48ms/step
1/1                     0s 48ms/step
1/1                     0s 49ms/step
1/1                     0s 51ms/step
```

```
1/1                  0s 44ms/step
1/1                  0s 41ms/step
1/1                  0s 55ms/step
1/1                  0s 40ms/step
1/1                  0s 42ms/step
1/1                  0s 38ms/step
1/1                  0s 38ms/step
1/1                  0s 52ms/step
1/1                  0s 41ms/step
1/1                  0s 50ms/step
1/1                  0s 45ms/step
1/1                  0s 52ms/step
1/1                  0s 42ms/step
1/1                  0s 41ms/step
1/1                  0s 49ms/step
1/1                  0s 44ms/step
1/1                  0s 51ms/step
1/1                  0s 54ms/step
1/1                  0s 69ms/step
1/1                  0s 49ms/step
1/1                  0s 57ms/step
1/1                  0s 55ms/step
1/1                  0s 34ms/step
1/1                  0s 34ms/step
1/1                  0s 34ms/step
1/1                  0s 35ms/step
1/1                  0s 35ms/step
1/1                  0s 35ms/step
1/1                  0s 40ms/step
1/1                  0s 107ms/step
1/1                  0s 153ms/step
1/1                  0s 85ms/step
1/1                  0s 32ms/step
1/1                  0s 33ms/step
1/1                  0s 43ms/step
1/1                  0s 33ms/step
1/1                  0s 31ms/step
1/1                  0s 31ms/step
1/1                  0s 32ms/step
1/1                  0s 33ms/step
1/1                  0s 36ms/step
1/1                  0s 37ms/step
1/1                  0s 33ms/step
1/1                  0s 32ms/step
1/1                  0s 33ms/step
1/1                  0s 33ms/step
1/1                  0s 32ms/step
1/1                  0s 36ms/step
```

```
1/1                   0s 32ms/step
1/1                   0s 33ms/step
1/1                   0s 33ms/step
1/1                   0s 30ms/step
1/1                   0s 30ms/step
1/1                   0s 33ms/step
1/1                   0s 30ms/step
1/1                   0s 34ms/step
1/1                   0s 32ms/step
1/1                   0s 35ms/step
1/1                   0s 32ms/step
1/1                   0s 35ms/step
1/1                   0s 33ms/step
1/1                   0s 31ms/step
1/1                   0s 32ms/step
1/1                   0s 32ms/step
1/1                   0s 33ms/step
1/1                   0s 32ms/step
1/1                   0s 34ms/step
1/1                   0s 34ms/step
1/1                   0s 32ms/step
1/1                   0s 30ms/step
1/1                   0s 39ms/step
1/1                   0s 32ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 33ms/step
1/1                   0s 36ms/step
1/1                   0s 33ms/step
1/1                   0s 31ms/step
1/1                   0s 30ms/step
1/1                   0s 30ms/step
1/1                   0s 31ms/step
1/1                   0s 31ms/step
1/1                   0s 35ms/step
1/1                   0s 37ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 33ms/step
1/1                   0s 32ms/step
1/1                   0s 34ms/step
1/1                   0s 34ms/step
1/1                   0s 32ms/step
1/1                   0s 32ms/step
1/1                   0s 31ms/step
1/1                   0s 34ms/step
1/1                   0s 33ms/step
1/1                   0s 33ms/step
```

```
1/1                Os 35ms/step
1/1                Os 31ms/step
1/1                Os 32ms/step
1/1                Os 32ms/step
1/1                Os 32ms/step
1/1                Os 30ms/step
1/1                Os 31ms/step
1/1                Os 34ms/step
1/1                Os 31ms/step
1/1                Os 41ms/step
1/1                Os 33ms/step
1/1                Os 31ms/step
1/1                Os 33ms/step
1/1                Os 34ms/step
1/1                Os 34ms/step
1/1                Os 33ms/step
1/1                Os 32ms/step
1/1                Os 33ms/step
1/1                Os 30ms/step
1/1                Os 29ms/step
1/1                Os 30ms/step
1/1                Os 42ms/step
1/1                Os 36ms/step
1/1                Os 36ms/step
1/1                Os 33ms/step
1/1                Os 34ms/step
1/1                Os 33ms/step
1/1                Os 35ms/step
1/1                Os 32ms/step
1/1                Os 30ms/step
1/1                Os 30ms/step
1/1                Os 30ms/step
1/1                Os 32ms/step
1/1                Os 39ms/step
1/1                Os 48ms/step
1/1                Os 55ms/step
1/1                Os 42ms/step
1/1                Os 42ms/step
1/1                Os 43ms/step
1/1                Os 43ms/step
1/1                Os 39ms/step
1/1                Os 40ms/step
1/1                Os 40ms/step
1/1                Os 43ms/step
1/1                Os 46ms/step
1/1                Os 41ms/step
1/1                Os 51ms/step
1/1                Os 52ms/step
```

```
1/1                    0s 40ms/step
1/1                    0s 55ms/step
1/1                    0s 57ms/step
1/1                    0s 47ms/step
1/1                    0s 55ms/step
1/1                    0s 44ms/step
1/1                    0s 50ms/step
1/1                    0s 48ms/step
1/1                    0s 56ms/step
1/1                    0s 43ms/step
1/1                    0s 56ms/step
1/1                    0s 51ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 31ms/step
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 41ms/step
1/1                    0s 47ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 31ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
```

```
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 42ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 40ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 41ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 31ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 42ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
```

```
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 31ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 50ms/step
1/1                    0s 39ms/step
1/1                    0s 50ms/step
1/1                    0s 67ms/step
1/1                    0s 40ms/step
1/1                    0s 42ms/step
1/1                    0s 42ms/step
1/1                    0s 44ms/step
1/1                    0s 52ms/step
1/1                    0s 44ms/step
1/1                    0s 39ms/step
1/1                    0s 41ms/step
1/1                    0s 41ms/step
1/1                    0s 51ms/step
1/1                    0s 42ms/step
1/1                    0s 51ms/step
1/1                    0s 41ms/step
1/1                    0s 45ms/step
1/1                    0s 42ms/step
1/1                    0s 52ms/step
1/1                    0s 39ms/step
1/1                    0s 42ms/step
1/1                    0s 42ms/step
1/1                    0s 45ms/step
1/1                    0s 48ms/step
1/1                    0s 41ms/step
1/1                    0s 42ms/step
1/1                    0s 57ms/step
```

```
1/1                    0s 49ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 31ms/step
1/1                    0s 31ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 31ms/step
1/1                    0s 43ms/step
```

```
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 39ms/step
1/1                    0s 35ms/step
1/1                    0s 31ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 36ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 32ms/step
1/1                    0s 33ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
```

```
1/1                     0s 34ms/step
1/1                     0s 36ms/step
1/1                     0s 35ms/step
1/1                     0s 40ms/step
1/1                     0s 34ms/step
1/1                     0s 33ms/step
1/1                     0s 35ms/step
1/1                     0s 35ms/step
1/1                     0s 34ms/step
1/1                     0s 31ms/step
1/1                     0s 33ms/step
1/1                     0s 35ms/step
1/1                     0s 51ms/step
1/1                     0s 49ms/step
1/1                     0s 63ms/step
1/1                     0s 53ms/step
1/1                     0s 71ms/step
1/1                     0s 53ms/step
1/1                     0s 39ms/step
1/1                     0s 54ms/step
1/1                     0s 42ms/step
1/1                     0s 38ms/step
1/1                     0s 40ms/step
1/1                     0s 56ms/step
1/1                     0s 41ms/step
1/1                     0s 59ms/step
1/1                     0s 49ms/step
1/1                     0s 44ms/step
1/1                     0s 46ms/step
1/1                     0s 57ms/step
1/1                     0s 39ms/step
1/1                     0s 39ms/step
1/1                     0s 44ms/step
1/1                     0s 47ms/step
1/1                     0s 50ms/step
1/1                     0s 47ms/step
1/1                     0s 41ms/step
1/1                     0s 48ms/step
1/1                     0s 55ms/step
1/1                     0s 61ms/step
1/1                     0s 51ms/step
1/1                     0s 35ms/step
1/1                     0s 32ms/step
1/1                     0s 32ms/step
1/1                     0s 32ms/step
1/1                     0s 34ms/step
1/1                     0s 34ms/step
1/1                     0s 37ms/step
```

```
1/1                   0s 34ms/step
1/1                   0s 34ms/step
1/1                   0s 34ms/step
1/1                   0s 36ms/step
1/1                   0s 36ms/step
1/1                   0s 35ms/step
1/1                   0s 33ms/step
1/1                   0s 32ms/step
1/1                   0s 31ms/step
1/1                   0s 32ms/step
1/1                   0s 34ms/step
1/1                   0s 36ms/step
1/1                   0s 34ms/step
1/1                   0s 40ms/step
1/1                   0s 39ms/step
1/1                   0s 41ms/step
1/1                   0s 36ms/step
1/1                   0s 41ms/step
1/1                   0s 38ms/step
1/1                   0s 35ms/step
1/1                   0s 36ms/step
1/1                   0s 38ms/step
1/1                   0s 36ms/step
1/1                   0s 37ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 33ms/step
1/1                   0s 37ms/step
1/1                   0s 35ms/step
1/1                   0s 37ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 32ms/step
1/1                   0s 34ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
1/1                   0s 36ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
1/1                   0s 35ms/step
1/1                   0s 51ms/step
1/1                   0s 35ms/step
1/1                   0s 37ms/step
1/1                   0s 39ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
```

```
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 43ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 32ms/step
1/1                    0s 46ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 54ms/step
```

```
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 51ms/step
1/1                    0s 57ms/step
1/1                    0s 59ms/step
1/1                    0s 45ms/step
1/1                    0s 43ms/step
1/1                    0s 50ms/step
1/1                    0s 41ms/step
1/1                    0s 39ms/step
1/1                    0s 45ms/step
1/1                    0s 45ms/step
1/1                    0s 48ms/step
1/1                    0s 44ms/step
1/1                    0s 53ms/step
1/1                    0s 50ms/step
1/1                    0s 55ms/step
1/1                    0s 46ms/step
1/1                    0s 45ms/step
1/1                    0s 39ms/step
1/1                    0s 55ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 49ms/step
1/1                    0s 58ms/step
1/1                    0s 52ms/step
1/1                    0s 51ms/step
1/1                    0s 46ms/step
1/1                    0s 48ms/step
1/1                    0s 44ms/step
1/1                    0s 48ms/step
1/1                    0s 44ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 50ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 36ms/step
```

```
1/1                  0s 37ms/step
1/1                  0s 36ms/step
1/1                  0s 33ms/step
1/1                  0s 36ms/step
1/1                  0s 34ms/step
1/1                  0s 35ms/step
1/1                  0s 35ms/step
1/1                  0s 35ms/step
1/1                  0s 33ms/step
1/1                  0s 34ms/step
1/1                  0s 45ms/step
1/1                  0s 36ms/step
1/1                  0s 39ms/step
1/1                  0s 34ms/step
1/1                  0s 35ms/step
1/1                  0s 36ms/step
1/1                  0s 36ms/step
1/1                  0s 35ms/step
1/1                  0s 32ms/step
1/1                  0s 32ms/step
1/1                  0s 33ms/step
1/1                  0s 35ms/step
1/1                  0s 33ms/step
1/1                  0s 36ms/step
1/1                  0s 36ms/step
1/1                  0s 34ms/step
1/1                  0s 39ms/step
1/1                  0s 36ms/step
1/1                  0s 35ms/step
1/1                  0s 34ms/step
1/1                  0s 37ms/step
1/1                  0s 33ms/step
1/1                  0s 33ms/step
1/1                  0s 32ms/step
1/1                  0s 31ms/step
1/1                  0s 34ms/step
1/1                  0s 37ms/step
1/1                  0s 38ms/step
1/1                  0s 36ms/step
1/1                  0s 36ms/step
1/1                  0s 35ms/step
1/1                  0s 33ms/step
1/1                  0s 34ms/step
1/1                  0s 33ms/step
1/1                  0s 33ms/step
1/1                  0s 32ms/step
1/1                  0s 33ms/step
1/1                  0s 35ms/step
```

```
1/1              0s 37ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 33ms/step
1/1              0s 33ms/step
1/1              0s 37ms/step
1/1              0s 32ms/step
1/1              0s 34ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 41ms/step
1/1              0s 34ms/step
1/1              0s 44ms/step
1/1              0s 36ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 33ms/step
1/1              0s 34ms/step
1/1              0s 32ms/step
1/1              0s 37ms/step
1/1              0s 35ms/step
1/1              0s 37ms/step
1/1              0s 34ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 34ms/step
1/1              0s 37ms/step
1/1              0s 41ms/step
1/1              0s 38ms/step
1/1              0s 33ms/step
1/1              0s 33ms/step
1/1              0s 38ms/step
1/1              0s 33ms/step
1/1              0s 37ms/step
1/1              0s 35ms/step
1/1              0s 41ms/step
1/1              0s 38ms/step
1/1              0s 34ms/step
1/1              0s 33ms/step
1/1              0s 44ms/step
1/1              0s 42ms/step
1/1              0s 42ms/step
1/1              0s 42ms/step
1/1              0s 53ms/step
1/1              0s 45ms/step
1/1              0s 52ms/step
```

```
1/1                  0s 41ms/step
1/1                  0s 43ms/step
1/1                  0s 42ms/step
1/1                  0s 44ms/step
1/1                  0s 49ms/step
1/1                  0s 49ms/step
1/1                  0s 42ms/step
1/1                  0s 63ms/step
1/1                  0s 46ms/step
1/1                  0s 45ms/step
1/1                  0s 51ms/step
1/1                  0s 42ms/step
1/1                  0s 40ms/step
1/1                  0s 54ms/step
1/1                  0s 56ms/step
1/1                  0s 52ms/step
1/1                  0s 51ms/step
1/1                  0s 54ms/step
1/1                  0s 44ms/step
1/1                  0s 46ms/step
1/1                  0s 50ms/step
1/1                  0s 43ms/step
1/1                  0s 33ms/step
1/1                  0s 40ms/step
1/1                  0s 36ms/step
1/1                  0s 34ms/step
1/1                  0s 46ms/step
1/1                  0s 34ms/step
1/1                  0s 34ms/step
1/1                  0s 34ms/step
1/1                  0s 34ms/step
1/1                  0s 33ms/step
1/1                  0s 36ms/step
1/1                  0s 32ms/step
1/1                  0s 34ms/step
1/1                  0s 37ms/step
1/1                  0s 36ms/step
1/1                  0s 36ms/step
1/1                  0s 36ms/step
1/1                  0s 38ms/step
1/1                  0s 37ms/step
1/1                  0s 35ms/step
1/1                  0s 37ms/step
1/1                  0s 36ms/step
1/1                  0s 42ms/step
1/1                  0s 33ms/step
1/1                  0s 34ms/step
1/1                  0s 44ms/step
```

```
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 38ms/step
1/1                    0s 34ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 32ms/step
1/1                    0s 43ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 42ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
```

```
1/1                   0s 36ms/step
1/1                   0s 44ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
1/1                   0s 33ms/step
1/1                   0s 32ms/step
1/1                   0s 36ms/step
1/1                   0s 36ms/step
1/1                   0s 37ms/step
1/1                   0s 33ms/step
1/1                   0s 35ms/step
1/1                   0s 34ms/step
1/1                   0s 40ms/step
1/1                   0s 35ms/step
1/1                   0s 36ms/step
1/1                   0s 37ms/step
1/1                   0s 36ms/step
1/1                   0s 35ms/step
1/1                   0s 41ms/step
1/1                   0s 34ms/step
1/1                   0s 35ms/step
1/1                   0s 36ms/step
1/1                   0s 36ms/step
1/1                   0s 35ms/step
1/1                   0s 36ms/step
1/1                   0s 37ms/step
1/1                   0s 35ms/step
1/1                   0s 37ms/step
1/1                   0s 35ms/step
1/1                   0s 35ms/step
1/1                   0s 36ms/step
1/1                   0s 35ms/step
1/1                   0s 43ms/step
1/1                   0s 39ms/step
1/1                   0s 49ms/step
1/1                   0s 62ms/step
1/1                   0s 43ms/step
1/1                   0s 54ms/step
1/1                   0s 43ms/step
1/1                   0s 58ms/step
1/1                   0s 56ms/step
1/1                   0s 51ms/step
1/1                   0s 44ms/step
1/1                   0s 57ms/step
1/1                   0s 54ms/step
1/1                   0s 61ms/step
1/1                   0s 56ms/step
```

```
1/1                0s 43ms/step
1/1                0s 58ms/step
1/1                0s 48ms/step
1/1                0s 58ms/step
1/1                0s 51ms/step
1/1                0s 53ms/step
1/1                0s 48ms/step
1/1                0s 50ms/step
1/1                0s 52ms/step
1/1                0s 57ms/step
1/1                0s 44ms/step
1/1                0s 60ms/step
1/1                0s 45ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 40ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 34ms/step
1/1                0s 38ms/step
1/1                0s 33ms/step
1/1                0s 33ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 35ms/step
1/1                0s 33ms/step
1/1                0s 33ms/step
1/1                0s 39ms/step
1/1                0s 38ms/step
1/1                0s 39ms/step
1/1                0s 35ms/step
1/1                0s 40ms/step
1/1                0s 46ms/step
1/1                0s 39ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 33ms/step
1/1                0s 35ms/step
1/1                0s 33ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
```

```
1/1                0s 37ms/step
1/1                0s 39ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 37ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
1/1                0s 38ms/step
1/1                0s 33ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
1/1                0s 35ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
1/1                0s 33ms/step
1/1                0s 35ms/step
1/1                0s 42ms/step
1/1                0s 34ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
1/1                0s 38ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 33ms/step
1/1                0s 35ms/step
1/1                0s 35ms/step
1/1                0s 34ms/step
1/1                0s 47ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 35ms/step
1/1                0s 33ms/step
1/1                0s 36ms/step
1/1                0s 32ms/step
1/1                0s 33ms/step
1/1                0s 38ms/step
```

```
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 34ms/step
1/1                0s 33ms/step
1/1                0s 39ms/step
1/1                0s 34ms/step
1/1                0s 33ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 43ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 38ms/step
1/1                0s 52ms/step
1/1                0s 39ms/step
1/1                0s 49ms/step
1/1                0s 41ms/step
1/1                0s 68ms/step
1/1                0s 49ms/step
1/1                0s 58ms/step
1/1                0s 47ms/step
1/1                0s 56ms/step
1/1                0s 43ms/step
1/1                0s 59ms/step
1/1                0s 49ms/step
1/1                0s 44ms/step
1/1                0s 43ms/step
1/1                0s 51ms/step
1/1                0s 45ms/step
1/1                0s 52ms/step
1/1                0s 41ms/step
1/1                0s 54ms/step
1/1                0s 52ms/step
1/1                0s 46ms/step
1/1                0s 63ms/step
1/1                0s 48ms/step
1/1                0s 49ms/step
1/1                0s 43ms/step
1/1                0s 63ms/step
1/1                0s 41ms/step
1/1                0s 58ms/step
1/1                0s 46ms/step
```

```
1/1                0s 37ms/step
1/1                0s 40ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 38ms/step
1/1                0s 35ms/step
1/1                0s 34ms/step
1/1                0s 32ms/step
1/1                0s 36ms/step
1/1                0s 39ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 33ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 32ms/step
1/1                0s 33ms/step
1/1                0s 41ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 38ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 33ms/step
1/1                0s 32ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 45ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 33ms/step
1/1                0s 34ms/step
1/1                0s 33ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
```

```
1/1                    0s 41ms/step
1/1                    0s 45ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 39ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 39ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 33ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 52ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 45ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 35ms/step
```

```
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 41ms/step
1/1                    0s 48ms/step
1/1                    0s 56ms/step
1/1                    0s 52ms/step
1/1                    0s 43ms/step
1/1                    0s 50ms/step
1/1                    0s 50ms/step
1/1                    0s 59ms/step
1/1                    0s 53ms/step
1/1                    0s 56ms/step
1/1                    0s 57ms/step
1/1                    0s 50ms/step
1/1                    0s 54ms/step
1/1                    0s 47ms/step
1/1                    0s 41ms/step
1/1                    0s 40ms/step
1/1                    0s 50ms/step
1/1                    0s 54ms/step
1/1                    0s 41ms/step
1/1                    0s 43ms/step
1/1                    0s 45ms/step
1/1                    0s 46ms/step
1/1                    0s 48ms/step
1/1                    0s 43ms/step
1/1                    0s 54ms/step
1/1                    0s 60ms/step
1/1                    0s 55ms/step
1/1                    0s 48ms/step
1/1                    0s 34ms/step
1/1                    0s 206ms/step
1/1                    0s 164ms/step
1/1                    0s 62ms/step
1/1                    0s 74ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 181ms/step
1/1                    0s 129ms/step
1/1                    0s 49ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
```

```
1/1                    0s 41ms/step
1/1                    0s 35ms/step
1/1                    0s 43ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 41ms/step
1/1                    0s 45ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 40ms/step
1/1                    0s 39ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 39ms/step
1/1                    0s 33ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 41ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
```

```
1/1                  0s 39ms/step
1/1                  0s 37ms/step
1/1                  0s 36ms/step
1/1                  0s 38ms/step
1/1                  0s 37ms/step
1/1                  0s 37ms/step
1/1                  0s 35ms/step
1/1                  0s 34ms/step
1/1                  0s 39ms/step
1/1                  0s 36ms/step
1/1                  0s 41ms/step
1/1                  0s 37ms/step
1/1                  0s 40ms/step
1/1                  0s 37ms/step
1/1                  0s 37ms/step
1/1                  0s 39ms/step
1/1                  0s 37ms/step
1/1                  0s 38ms/step
1/1                  0s 41ms/step
1/1                  0s 36ms/step
1/1                  0s 34ms/step
1/1                  0s 36ms/step
1/1                  0s 35ms/step
1/1                  0s 35ms/step
1/1                  0s 41ms/step
1/1                  0s 42ms/step
1/1                  0s 38ms/step
1/1                  0s 37ms/step
1/1                  0s 53ms/step
1/1                  0s 44ms/step
1/1                  0s 43ms/step
1/1                  0s 42ms/step
1/1                  0s 42ms/step
1/1                  0s 47ms/step
1/1                  0s 62ms/step
1/1                  0s 53ms/step
1/1                  0s 61ms/step
1/1                  0s 63ms/step
1/1                  0s 43ms/step
1/1                  0s 45ms/step
1/1                  0s 49ms/step
1/1                  0s 43ms/step
1/1                  0s 44ms/step
1/1                  0s 53ms/step
1/1                  0s 62ms/step
1/1                  0s 65ms/step
1/1                  0s 45ms/step
1/1                  0s 45ms/step
```

```
1/1                    0s 45ms/step
1/1                    0s 51ms/step
1/1                    0s 41ms/step
1/1                    0s 63ms/step
1/1                    0s 46ms/step
1/1                    0s 63ms/step
1/1                    0s 45ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 34ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 48ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 42ms/step
1/1                    0s 39ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 41ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 37ms/step
1/1                    0s 44ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
```

```
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 39ms/step
1/1                0s 33ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 39ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 40ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
1/1                0s 35ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 41ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 38ms/step
1/1                0s 39ms/step
1/1                0s 36ms/step
1/1                0s 34ms/step
1/1                0s 47ms/step
1/1                0s 36ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 34ms/step
```

```
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 36ms/step
1/1                    0s 39ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 39ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 50ms/step
1/1                    0s 43ms/step
1/1                    0s 46ms/step
1/1                    0s 43ms/step
1/1                    0s 41ms/step
1/1                    0s 43ms/step
1/1                    0s 53ms/step
1/1                    0s 47ms/step
1/1                    0s 55ms/step
1/1                    0s 51ms/step
1/1                    0s 54ms/step
1/1                    0s 50ms/step
1/1                    0s 42ms/step
1/1                    0s 50ms/step
1/1                    0s 68ms/step
1/1                    0s 58ms/step
1/1                    0s 55ms/step
1/1                    0s 52ms/step
1/1                    0s 58ms/step
1/1                    0s 46ms/step
1/1                    0s 67ms/step
1/1                    0s 48ms/step
1/1                    0s 46ms/step
1/1                    0s 51ms/step
1/1                    0s 43ms/step
1/1                    0s 53ms/step
1/1                    0s 50ms/step
1/1                    0s 47ms/step
1/1                    0s 49ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
```

```
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 49ms/step
1/1                0s 34ms/step
1/1                0s 38ms/step
1/1                0s 34ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 38ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 43ms/step
1/1                0s 34ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
1/1                0s 39ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 37ms/step
1/1                0s 34ms/step
1/1                0s 34ms/step
1/1                0s 38ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 40ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 39ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 41ms/step
1/1                0s 35ms/step
1/1                0s 36ms/step
```

```
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 41ms/step
1/1                    0s 39ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 42ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 41ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 41ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 40ms/step
1/1                    0s 40ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 59ms/step
```

```
1/1                    0s 43ms/step
1/1                    0s 45ms/step
1/1                    0s 54ms/step
1/1                    0s 45ms/step
1/1                    0s 53ms/step
1/1                    0s 59ms/step
1/1                    0s 55ms/step
1/1                    0s 66ms/step
1/1                    0s 55ms/step
1/1                    0s 40ms/step
1/1                    0s 45ms/step
1/1                    0s 46ms/step
1/1                    0s 44ms/step
1/1                    0s 41ms/step
1/1                    0s 45ms/step
1/1                    0s 45ms/step
1/1                    0s 43ms/step
1/1                    0s 45ms/step
1/1                    0s 48ms/step
1/1                    0s 41ms/step
1/1                    0s 45ms/step
1/1                    0s 43ms/step
1/1                    0s 48ms/step
1/1                    0s 47ms/step
1/1                    0s 56ms/step
1/1                    0s 50ms/step
1/1                    0s 50ms/step
1/1                    0s 48ms/step
1/1                    0s 53ms/step
1/1                    0s 47ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 40ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
```

```
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 40ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 46ms/step
1/1                    0s 36ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 40ms/step
1/1                    0s 38ms/step
1/1                    0s 42ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 43ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
```

```
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 39ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 39ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 41ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 44ms/step
1/1                    0s 34ms/step
1/1                    0s 46ms/step
1/1                    0s 54ms/step
1/1                    0s 42ms/step
1/1                    0s 54ms/step
1/1                    0s 43ms/step
1/1                    0s 56ms/step
1/1                    0s 55ms/step
1/1                    0s 78ms/step
1/1                    0s 42ms/step
1/1                    0s 50ms/step
1/1                    0s 44ms/step
1/1                    0s 47ms/step
1/1                    0s 49ms/step
1/1                    0s 40ms/step
1/1                    0s 40ms/step
```

```
1/1                    0s 57ms/step
1/1                    0s 53ms/step
1/1                    0s 76ms/step
1/1                    0s 60ms/step
1/1                    0s 62ms/step
1/1                    0s 49ms/step
1/1                    0s 46ms/step
1/1                    0s 49ms/step
1/1                    0s 49ms/step
1/1                    0s 55ms/step
1/1                    0s 41ms/step
1/1                    0s 46ms/step
1/1                    0s 43ms/step
1/1                    0s 55ms/step
1/1                    0s 46ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 46ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 33ms/step
```

```
1/1                    0s 36ms/step
1/1                    0s 44ms/step
1/1                    0s 39ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 50ms/step
1/1                    0s 42ms/step
1/1                    0s 40ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 45ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 40ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 33ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 34ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
```

```
1/1                0s 35ms/step
1/1                0s 37ms/step
1/1                0s 38ms/step
1/1                0s 38ms/step
1/1                0s 49ms/step
1/1                0s 38ms/step
1/1                0s 37ms/step
1/1                0s 35ms/step
1/1                0s 38ms/step
1/1                0s 35ms/step
1/1                0s 33ms/step
1/1                0s 33ms/step
1/1                0s 36ms/step
1/1                0s 37ms/step
1/1                0s 41ms/step
1/1                0s 36ms/step
1/1                0s 36ms/step
1/1                0s 35ms/step
1/1                0s 46ms/step
1/1                0s 47ms/step
1/1                0s 43ms/step
1/1                0s 57ms/step
1/1                0s 45ms/step
1/1                0s 40ms/step
1/1                0s 58ms/step
1/1                0s 57ms/step
1/1                0s 50ms/step
1/1                0s 44ms/step
1/1                0s 44ms/step
1/1                0s 61ms/step
1/1                0s 53ms/step
1/1                0s 59ms/step
1/1                0s 58ms/step
1/1                0s 56ms/step
1/1                0s 48ms/step
1/1                0s 63ms/step
1/1                0s 63ms/step
1/1                0s 75ms/step
1/1                0s 61ms/step
1/1                0s 60ms/step
1/1                0s 67ms/step
1/1                0s 56ms/step
1/1                0s 55ms/step
1/1                0s 40ms/step
1/1                0s 59ms/step
1/1                0s 44ms/step
1/1                0s 70ms/step
1/1                0s 60ms/step
```

```
1/1                    0s 38ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 39ms/step
1/1                    0s 44ms/step
1/1                    0s 43ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 39ms/step
1/1                    0s 38ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 38ms/step
1/1                    0s 39ms/step
1/1                    0s 44ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 38ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 40ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
```

```
1/1                 0s 36ms/step
1/1                 0s 37ms/step
1/1                 0s 40ms/step
1/1                 0s 59ms/step
1/1                 0s 36ms/step
1/1                 0s 35ms/step
1/1                 0s 38ms/step
1/1                 0s 37ms/step
1/1                 0s 40ms/step
1/1                 0s 42ms/step
1/1                 0s 39ms/step
1/1                 0s 43ms/step
1/1                 0s 40ms/step
1/1                 0s 39ms/step
1/1                 0s 38ms/step
1/1                 0s 38ms/step
1/1                 0s 39ms/step
1/1                 0s 36ms/step
1/1                 0s 36ms/step
1/1                 0s 37ms/step
1/1                 0s 34ms/step
1/1                 0s 38ms/step
1/1                 0s 35ms/step
1/1                 0s 34ms/step
1/1                 0s 38ms/step
1/1                 0s 35ms/step
1/1                 0s 36ms/step
1/1                 0s 36ms/step
1/1                 0s 38ms/step
1/1                 0s 35ms/step
1/1                 0s 37ms/step
1/1                 0s 37ms/step
1/1                 0s 35ms/step
1/1                 0s 35ms/step
1/1                 0s 34ms/step
1/1                 0s 34ms/step
1/1                 0s 38ms/step
1/1                 0s 37ms/step
1/1                 0s 35ms/step
1/1                 0s 42ms/step
1/1                 0s 36ms/step
1/1                 0s 37ms/step
1/1                 0s 43ms/step
1/1                 0s 35ms/step
1/1                 0s 39ms/step
1/1                 0s 36ms/step
1/1                 0s 34ms/step
1/1                 0s 35ms/step
```

```
1/1                    0s 40ms/step
1/1                    0s 36ms/step
1/1                    0s 36ms/step
1/1                    0s 51ms/step
1/1                    0s 49ms/step
1/1                    0s 45ms/step
1/1                    0s 52ms/step
1/1                    0s 41ms/step
1/1                    0s 44ms/step
1/1                    0s 41ms/step
1/1                    0s 39ms/step
1/1                    0s 54ms/step
1/1                    0s 43ms/step
1/1                    0s 42ms/step
1/1                    0s 41ms/step
1/1                    0s 41ms/step
1/1                    0s 44ms/step
1/1                    0s 60ms/step
1/1                    0s 56ms/step
1/1                    0s 41ms/step
1/1                    0s 47ms/step
1/1                    0s 53ms/step
1/1                    0s 48ms/step
1/1                    0s 48ms/step
1/1                    0s 43ms/step
1/1                    0s 68ms/step
1/1                    0s 61ms/step
1/1                    0s 57ms/step
1/1                    0s 51ms/step
1/1                    0s 54ms/step
1/1                    0s 54ms/step
1/1                    0s 48ms/step
1/1                    0s 47ms/step
1/1                    0s 46ms/step
1/1                    0s 70ms/step
1/1                    0s 33ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
1/1                    0s 40ms/step
1/1                    0s 37ms/step
1/1                    0s 40ms/step
1/1                    0s 38ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 36ms/step
1/1                    0s 35ms/step
1/1                    0s 35ms/step
```

```
1/1                     0s 35ms/step
1/1                     0s 39ms/step
1/1                     0s 39ms/step
1/1                     0s 41ms/step
1/1                     0s 36ms/step
1/1                     0s 40ms/step
1/1                     0s 37ms/step
1/1                     0s 36ms/step
1/1                     0s 41ms/step
1/1                     0s 34ms/step
1/1                     0s 34ms/step
1/1                     0s 36ms/step
1/1                     0s 37ms/step
1/1                     0s 45ms/step
1/1                     0s 41ms/step
1/1                     0s 38ms/step
1/1                     0s 36ms/step
1/1                     0s 39ms/step
1/1                     0s 38ms/step
1/1                     0s 36ms/step
1/1                     0s 41ms/step
1/1                     0s 34ms/step
1/1                     0s 35ms/step
1/1                     0s 33ms/step
1/1                     0s 37ms/step
1/1                     0s 39ms/step
1/1                     0s 41ms/step
1/1                     0s 37ms/step
1/1                     0s 35ms/step
1/1                     0s 38ms/step
1/1                     0s 34ms/step
1/1                     0s 39ms/step
1/1                     0s 37ms/step
1/1                     0s 36ms/step
1/1                     0s 34ms/step
1/1                     0s 35ms/step
1/1                     0s 37ms/step
1/1                     0s 39ms/step
1/1                     0s 38ms/step
1/1                     0s 42ms/step
1/1                     0s 35ms/step
1/1                     0s 41ms/step
1/1                     0s 36ms/step
1/1                     0s 34ms/step
1/1                     0s 43ms/step
1/1                     0s 35ms/step
1/1                     0s 35ms/step
1/1                     0s 38ms/step
```

```
1/1          0s 35ms/step
1/1          0s 45ms/step
1/1          0s 42ms/step
1/1          0s 37ms/step
1/1          0s 36ms/step
1/1          0s 35ms/step
1/1          0s 37ms/step
1/1          0s 38ms/step
1/1          0s 35ms/step
1/1          0s 34ms/step
1/1          0s 36ms/step
1/1          0s 34ms/step
1/1          0s 36ms/step
1/1          0s 38ms/step
1/1          0s 40ms/step
1/1          0s 36ms/step
1/1          0s 37ms/step
1/1          0s 36ms/step
1/1          0s 37ms/step
1/1          0s 40ms/step
1/1          0s 35ms/step
1/1          0s 35ms/step
1/1          0s 35ms/step
1/1          0s 36ms/step
1/1          0s 35ms/step
1/1          0s 44ms/step
1/1          0s 36ms/step
1/1          0s 35ms/step
1/1          0s 38ms/step
1/1          0s 35ms/step
1/1          0s 34ms/step
1/1          0s 37ms/step
1/1          0s 41ms/step
1/1          0s 35ms/step
1/1          0s 36ms/step
1/1          0s 36ms/step
1/1          0s 44ms/step
1/1          0s 60ms/step
1/1          0s 70ms/step
1/1          0s 56ms/step
1/1          0s 64ms/step
1/1          0s 52ms/step
1/1          0s 59ms/step
1/1          0s 49ms/step
1/1          0s 69ms/step
1/1          0s 47ms/step
1/1          0s 40ms/step
1/1          0s 51ms/step
```

```
1/1                    0s 47ms/step
1/1                    0s 43ms/step
1/1                    0s 48ms/step
1/1                    0s 55ms/step
1/1                    0s 50ms/step
1/1                    0s 43ms/step
1/1                    0s 45ms/step
1/1                    0s 43ms/step
1/1                    0s 44ms/step
1/1                    0s 40ms/step
1/1                    0s 46ms/step
1/1                    0s 58ms/step
1/1                    0s 52ms/step
1/1                    0s 66ms/step
1/1                    0s 43ms/step
1/1                    0s 74ms/step
1/1                    0s 59ms/step
1/1                    0s 56ms/step
1/1                    0s 51ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 47ms/step
1/1                    0s 32ms/step
1/1                    0s 34ms/step
1/1                    0s 40ms/step
1/1                    0s 38ms/step
1/1                    0s 41ms/step
1/1                    0s 37ms/step
1/1                    0s 39ms/step
1/1                    0s 35ms/step
1/1                    0s 44ms/step
1/1                    0s 35ms/step
1/1                    0s 37ms/step
1/1                    0s 35ms/step
1/1                    0s 34ms/step
1/1                    0s 35ms/step
1/1                    0s 38ms/step
1/1                    0s 36ms/step
1/1                    0s 39ms/step
1/1                    0s 43ms/step
1/1                    0s 36ms/step
1/1                    0s 37ms/step
1/1                    0s 44ms/step
1/1                    0s 38ms/step
1/1                    0s 34ms/step
1/1                    0s 37ms/step
1/1                    0s 34ms/step
1/1                    0s 33ms/step
```

```
1/1              0s 40ms/step
1/1              0s 37ms/step
1/1              0s 38ms/step
1/1              0s 40ms/step
1/1              0s 37ms/step
1/1              0s 39ms/step
1/1              0s 37ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 35ms/step
1/1              0s 38ms/step
1/1              0s 38ms/step
1/1              0s 36ms/step
1/1              0s 38ms/step
1/1              0s 35ms/step
1/1              0s 38ms/step
1/1              0s 37ms/step
1/1              0s 37ms/step
1/1              0s 34ms/step
1/1              0s 35ms/step
1/1              0s 37ms/step
1/1              0s 34ms/step
1/1              0s 34ms/step
1/1              0s 38ms/step
1/1              0s 46ms/step
1/1              0s 37ms/step
```

```
[ ]:  # We can save the word2vec dictionary to reuse it later.
      #new_word2vec.to_csv('word2vec_MutualFundSummary_train.txt')
```

```
[44]:  type(word2vec)
```

[44]: dict

```
[ ]:  #import pickle
      #file_path_pickle = 'word2vec_model.pkl'
      #print(f"Saving dictionary to {file_path_pickle} using pickle...")
      #with open(file_path_pickle, 'wb') as f: # 'wb' means write binary
      #    pickle.dump(word2vec, f)
      #print("Saving complete.")
```

```
      Saving dictionary to word2vec_model.pkl using pickle…
      Saving complete.
```

```
[ ]:  #word2vec = load_word2vec("word2vec_perso.txt")
```

```
[45]: # for a given word, output the n closer words in the word2vec maping.
      def get_n_closer(w, n, word2vec):
          vect = word2vec[w]
          distances_dict = {k: cosine(v, vect) for k, v in word2vec.items()}

          closer_words = []
          for _ in range(n):
              min_key = min(distances_dict.keys(), key=lambda k: distances_dict[k])
              closer_words.append(min_key)
              del distances_dict[min_key]
          return closer_words
```

```
[49]: words_neighbors_1 = get_n_closer('risk', 10, word2vec)
      print('words close to risk : ' +str(', '.join(words_neighbors_1)))
      #words_neighbors_2 = get_n_closer('equiderivativety', 10, word2vec)
      #print('words close to derivative : ' +str(', '.join(words_neighbors_2)))
```

words close to risk : risk, targets, slump, concerns, mortgage-, relative,
treated, unavailable, participation, although

```
[50]: #plot closer words
      # fit the TSNE dimenstion reduction model
      from sklearn.manifold import TSNE
      from sklearn.decomposition import PCA
      vectors = np.array([v for k, v in word2vec.items()])
      words = np.array([k for k, v in word2vec.items()])
      print(vectors.shape)

      tsne_model = TSNE(perplexity=40, n_components=2, init='pca', n_iter=2500,␣
       ↪random_state=23)
      vectors = tsne_model.fit_transform(vectors)

      print(vectors.shape)
```

(2699, 50)

/usr/local/lib/python3.11/dist-packages/sklearn/manifold/_t_sne.py:1164:
FutureWarning: 'n_iter' was renamed to 'max_iter' in version 1.5 and will be
removed in 1.7.
  warnings.warn(

(2699, 2)

```
[51]: reduced_word2vec = {words[i] : list(vectors[i]) for i in range (len(vectors))}
```

```
[52]: def create_neighbors_visualisation_data(words, num_neighbors, word2vec):
          words_neighbors = []
          for word in words :
```

```
        try :
            words_neighbors.append(get_n_closer(word, num_neighbors, word2vec))
        except :
            print(word +' not in word2vec')


    words_neighbors = np.concatenate(words_neighbors)
    vectors_neighbors = np.array([reduced_word2vec[word] for word in␣
  ↪words_neighbors])


    return vectors_neighbors[:,0], vectors_neighbors[:,0], words_neighbors
x, y, labels = create_neighbors_visualisation_data(['fee', 'swaps'], 10,␣
  ↪word2vec)
```

```
[53]: plt.figure(figsize=(16, 16))
for i in range(len(x)):
    plt.scatter(x[i],y[i])
    plt.annotate(labels[i],
        xy=(x[i], y[i]),
        xytext=(5, 2),
        textcoords='offset points',
        ha='right',
        va='bottom')
plt.show()
```

[ ]:
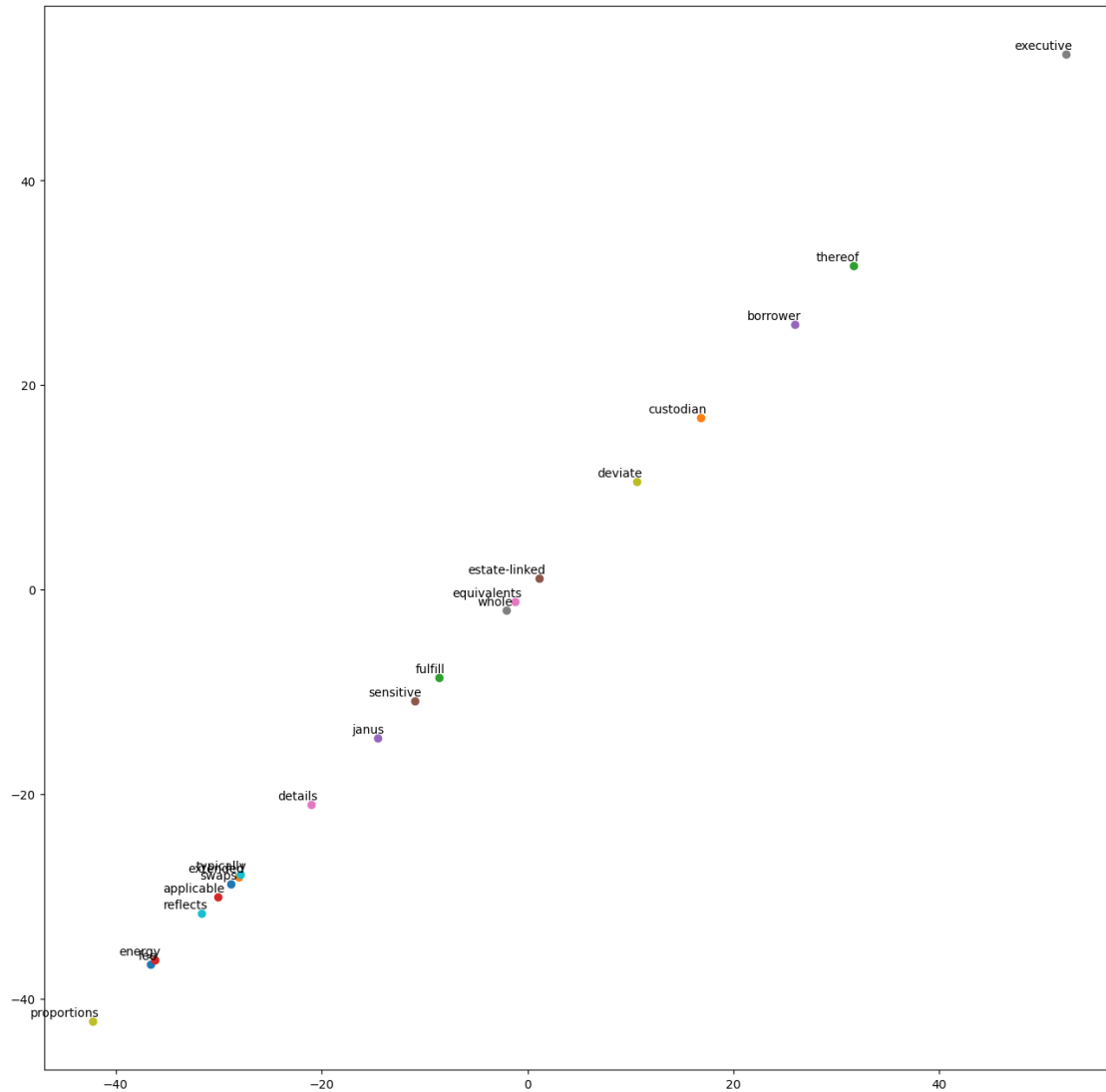
0.4 Create a knowledge base

```
[64]: key_words = [ 'income', 'interest', 'bond', 'long','short',
                'equity', 'stock','low', 'risk', 'risks','strategy',␣
      ↪'strategies','return', 'volatility',
                'fund','funds','balance','balanced', 'yield','position']
```

```
[55]: #print(f"\nLoading dictionary from {file_path_pickle} using pickle...")
      #with open('word2vec_model.pkl', 'rb') as f: # 'rb' means read binary
      #    word2vec = pickle.load(f)
      #print("Loading complete.")
```

```
[56]:  #word2vec = load_word2vec(new_word2vec)
```

```
[65]:  # Creates the knwoledge base by taking the num_neighbors closes neighbors of␣
       ↪each key_words in word2vec
       def create_knowledge_base(num_neighbors, word2vec, key_words):
           knowledge_base = set()
           out = display(progress(0, len(key_words)-1), display_id=True)
           for ii, key_word in enumerate(key_words) :
               knowledge_base.add(key_word)
               neighbors = []
               try :
                   neighbors = get_n_closer(key_word, num_neighbors, word2vec)
               except :
                   print(key_word + ' not in word2vec')

               knowledge_base.update(neighbors)

               out.update(progress(ii, len(key_words)-1))
           return knowledge_base
```

```
[66]:  #word2vec = load_word2vec(our_word2vec)
       knowledge_base = create_knowledge_base(5, word2vec, key_words)
```

<IPython.core.display.HTML object>

```
[67]:  print(knowledge_base)
```

{np.str_('operated'), np.str_('amt'), np.str_('targets'), np.str_('given'),
'position', 'yield', np.str_('fluctuate'), np.str_('mergers'),
np.str_('participants'), np.str_('obsolescence'), np.str_('terrorism'),
np.str_('tax-advantaged'), np.str_('available'), 'income', np.str_('advisers'),
np.str_('shareholder'), np.str_('assurances'), 'risk', np.str_('necessarily'),
'balance', np.str_('number'), np.str_('hearing'), np.str_('values'),
np.str_('less'), np.str_('unusually'), np.str_('jad'), np.str_('hypothetical'),
np.str_('screens'), np.str_('neutral'), np.str_('statement'), np.str_('equity-
income'), np.str_('region'), np.str_('modify'), np.str_('separate'),
np.str_('conflicts'), np.str_('frequency'), np.str_('investments'),
np.str_('experience'), 'stock', 'interest', np.str_('infrequent'),
np.str_('consumer'), np.str_('successful'), 'risks', np.str_('absent'),
np.str_('partners'), np.str_('contractually'), np.str_('merger'), 'volatility',
np.str_('outstanding'), 'funds', 'short', np.str_('regarding'),
np.str_('corporation'), np.str_('consisting'), np.str_('security'),
np.str_('log'), np.str_('savings'), np.str_('decreased'), np.str_('thereof'),
np.str_('mortgage-'), np.str_('suited'), np.str_('insurers'),
np.str_('admiral'), np.str_('dramatically'), np.str_('consequently'),
np.str_('repayment'), np.str_('across'), np.str_('technology-related'),
np.str_('sells'), 'strategies', np.str_('slump'), np.str_('residential'),

```
       np.str_('suite'), 'balanced', np.str_('australia'), 'strategy',
       np.str_('sales'), np.str_('disclosures'), np.str_('andrew'), np.str_('ensure'),
       np.str_('intermediaries'), np.str_('illustrates'), np.str_('frequently'),
       np.str_('concerns'), 'fund', 'low', np.str_('care'), np.str_('applies'),
       np.str_('mortgage-backed'), 'bond', np.str_('two'), 'long', 'equity',
       np.str_('appendix'), 'return', np.str_('reviewed')}
```

```
[63]: items_to_remove = {'balanced fund', 'fixed income', ' long only'}
      for item in items_to_remove:
          knowledge_base.discard(item)
      knowledge_base
```

```
[63]: {np.str_('absent'),
       np.str_('across'),
       np.str_('admiral'),
       np.str_('advisers'),
       np.str_('amt'),
       np.str_('andrew'),
       np.str_('appendix'),
       np.str_('applies'),
       np.str_('assurances'),
       np.str_('australia'),
       np.str_('available'),
       'balance',
       'balanced',
       'bond',
       np.str_('care'),
       np.str_('concerns'),
       np.str_('conflicts'),
       np.str_('consequently'),
       np.str_('consisting'),
       np.str_('consumer'),
       np.str_('contractually'),
       np.str_('corporation'),
       np.str_('decreased'),
       np.str_('disclosures'),
       np.str_('dramatically'),
       np.str_('ensure'),
       'equity',
       np.str_('equity-income'),
       np.str_('experience'),
       np.str_('fluctuate'),
       np.str_('frequency'),
       np.str_('frequently'),
       'fund',
       'funds',
       np.str_('given'),
```

```
np.str_('hearing'),
np.str_('hypothetical'),
np.str_('illustrates'),
'income',
np.str_('infrequent'),
np.str_('insurers'),
'interest',
np.str_('intermediaries'),
np.str_('investments'),
np.str_('jad'),
np.str_('less'),
np.str_('log'),
'long',
'low',
np.str_('merger'),
np.str_('mergers'),
np.str_('modify'),
np.str_('mortgage-'),
np.str_('mortgage-backed'),
np.str_('necessarily'),
np.str_('neutral'),
np.str_('number'),
np.str_('obsolescence'),
np.str_('operated'),
np.str_('outstanding'),
np.str_('participants'),
np.str_('partners'),
'position',
np.str_('regarding'),
np.str_('region'),
np.str_('repayment'),
np.str_('residential'),
'return',
np.str_('reviewed'),
'risk',
'risks',
np.str_('sales'),
np.str_('savings'),
np.str_('screens'),
np.str_('security'),
np.str_('sells'),
np.str_('separate'),
np.str_('shareholder'),
'short',
np.str_('slump'),
np.str_('statement'),
'stock',
```

```
         'strategies',
         'strategy',
         np.str_('successful'),
         np.str_('suite'),
         np.str_('suited'),
         np.str_('targets'),
         np.str_('tax-advantaged'),
         np.str_('technology-related'),
         np.str_('terrorism'),
         np.str_('thereof'),
         np.str_('two'),
         np.str_('unusually'),
         np.str_('values'),
         'volatility',
         'yield'}
```

[68]: `y_train[1]`

[68]: `'Janus Henderson Enterprise Portfolio'`

[69]: `df_summary[['fund_name','Prediction']]`

[69]:
```
                                           fund_name  Prediction
0                 American Funds College 2018 Fund          2.0
1                 American Funds College 2021 Fund          2.0
2                 American Funds College 2024 Fund          2.0
3                 American Funds College 2027 Fund          2.0
4                 American Funds College 2030 Fund          2.0
..                                             ...          ...
427    American Century Diversified Corporate Bond ETF      1.0
428    American Century Diversified Municipal Bond ETF      1.0
429  American Century Quality Diversified Internati…      0.0
430       American Century STOXX U.S. Quality Growth ETF    0.0
431        American Century STOXX U.S. Quality Value ETF    0.0

[432 rows x 2 columns]
```

[70]:
```python
# We create here the dataframe tha contains the summaries along with their␣
↪labels
df_extraction_train = pd.DataFrame({'fund_name' : y_train, 'summary':X_train})
df_label = df_summary[['fund_name','Prediction']]
df_main_train = df_label.merge(df_extraction_train, on='fund_name', how='inner')
df_main_train
```

[70]:
```
                                   fund_name  Prediction  \
0             American Funds College 2018 Fund          2.0
1             American Funds College 2021 Fund          2.0
```

```
2                     American Funds College 2024 Fund          2.0
3                     American Funds College 2033 Fund          2.0
4                     American Funds College 2036 Fund          2.0
..                                                    …            …
254                        American Century Growth ETF          0.0
255                         American Century Value ETF          0.0
256    American Century Diversified Corporate Bond ETF          1.0
257    American Century Diversified Municipal Bond ETF          1.0
258        American Century STOXX U.S. Quality Value ETF        0.0

                                                summary
0      American Funds College 2018 Fund\n\nInvestment…
1      American Funds College 2021 Fund\n\nInvestment…
2      American Funds College 2024 Fund\n\nInvestment…
3      American Funds College 2033 Fund\n\nInvestment…
4      American Funds College 2036 Fund\n\nInvestment…
..                                                   …
254    Fund Summary\nInvestment Objective\nThe fund s…
255    Fund Summary\nInvestment Objective\nThe fund s…
256    Fund Summary\nInvestment Objective\nThe fund s…
257    Fund Summary\nInvestment Objective\nThe fund s…
258    Fund Summary\nInvestment Objective\nThe fund s…

[259 rows x 3 columns]
```

[71]: `df_label.head()`

```
[71]:                          fund_name  Prediction
      0  American Funds College 2018 Fund          2.0
      1  American Funds College 2021 Fund          2.0
      2  American Funds College 2024 Fund          2.0
      3  American Funds College 2027 Fund          2.0
      4  American Funds College 2030 Fund          2.0
```

[72]: `df_extraction_train.head()`

```
[72]:                          fund_name  \
      0      MFS Maryland Municipal Bond Fund
      1  Janus Henderson Enterprise Portfolio
      2               Ivy Municipal Bond Fund
      3    Fidelity Advisor New Insights Fund
      4                     CAPITAL VALUE FUND


                                                summary
      0  MFS Maryland Municipal Bond Fund\n\n \n\nSumma…
      1  Janus Henderson Enterprise Fund\n(closed to ce…
      2  Ivy Municipal Bond Fund\n\nObjective\n\nTo see…
```

```
3  Fund Summary\n\nFund/Class:\nFidelity Advisor®…
4  Fund Summary \nInvestment Objective \nThe fund…
```

[73]:
```
# We create here the dataframe tha contains the summaries along with their
 ↪labels
df_extraction_val = pd.DataFrame({'fund_name' : y_val, 'summary':X_val})
df_label = df_summary[['fund_name','Prediction']]
df_main_val = df_label.merge(df_extraction_val, on='fund_name', how='inner')
df_main_val
```

[73]:
```
                                         fund_name  Prediction  \
0                  American Funds College 2027 Fund         2.0
1                  American Funds College 2030 Fund         2.0
2             American Funds College Enrollment Fund        1.0
3   American Funds U.S. Government Money Market Fund        1.0
4                              Global Balanced Fund         2.0
..                                             ...          ...
81            Anchor Tactical Credit Strategies Fund        1.0
82                 Federated MDT Small Cap Growth Fund      0.0
83                               S&P 500 Index Fund         0.0
84                           Managed Allocation Fund        2.0
85    American Century STOXX U.S. Quality Growth ETF        0.0

                                              summary
0    American Funds College 2027 Fund\n\nInvestment…
1    American Funds College 2030 Fund\n\nInvestment…
2    American Funds College Enrollment Fund\n\nInve…
3    Investment objective The investment objective …
4    Global Balanced Fund\n\nInvestment objectives …
..                                                …
81   FUND SUMMARY – ANCHOR TACTICAL CREDIT STRATEGI…
82   Fund Summary Information\nFederated MDT Small …
83   INVESTMENT OBJECTIVE\nThe USAA S&P 500 Index F…
84   INVESTMENT OBJECTIVE\nThe USAA Managed Allocat…
85   Fund Summary\nInvestment Objective\nThe fund s…

[86 rows x 3 columns]
```

[74]:
```
# We create here the dataframe tha contains the summaries along with their
 ↪labels
df_extraction_test = pd.DataFrame({'fund_name' : y_test, 'summary':X_test})
df_label = df_summary[['fund_name','Prediction']]
df_main_test = df_label.merge(df_extraction_test, on='fund_name', how='inner')
df_main_test
```

[74]:
```
                                         fund_name  Prediction  \
0                                         Bond Fund        1.0
```

```
1                         Corporate Bond Fund          1.0
2                           Global Bond Fund          1.0
3                Global Growth and Income Fund          2.0
4              Global Small Capitalization Fund          0.0
..                                        …            …
82              Federated MDT All Cap Core Fund          0.0
83                       Nasdaq-100 Index Fund          0.0
84                           Income Stock Fund          0.0
85   USAA MSCI International Value Momentum Blend I…    0.0
86   American Century Quality Diversified Internati…    0.0

                                         summary
0    Bond Fund\n\nInvestment objective The fund's i…
1    Investment objective The fund's investment obj…
2    Global Bond Fund\n\nInvestment objective The f…
3    Global Growth and Income Fund\n\nInvestment ob…
4    Global Small Capitalization Fund\n\nInvestment…
..                                              …
82   Fund Summary Information\nFederated MDT All Ca…
83   INVESTMENT OBJECTIVE\nThe USAA Nasdaq-100 Inde…
84   INVESTMENT OBJECTIVE\nThe USAA Income Stock Fu…
85   INVESTMENT OBJECTIVE\nThe USAA MSCI USA Value …
86   Fund Summary\nInvestment Objective\nThe fund s…

[87 rows x 3 columns]
```

[75]:
```python
df_main_train['summary'] = df_main_train['summary'].astype(str)
df_main_val['summary'] = df_main_val['summary'].astype(str)
df_main_test['summary'] = df_main_test['summary'].astype(str)
df_main_train
```

[75]:
```
                                     fund_name  Prediction  \
0                  American Funds College 2018 Fund        2.0
1                  American Funds College 2021 Fund        2.0
2                  American Funds College 2024 Fund        2.0
3                  American Funds College 2033 Fund        2.0
4                  American Funds College 2036 Fund        2.0
..                                           …          …
254                      American Century Growth ETF        0.0
255                       American Century Value ETF        0.0
256   American Century Diversified Corporate Bond ETF    1.0
257   American Century Diversified Municipal Bond ETF    1.0
258     American Century STOXX U.S. Quality Value ETF    0.0

                                         summary
0    American Funds College 2018 Fund\n\nInvestment…
1    American Funds College 2021 Fund\n\nInvestment…
```

```
2      American Funds College 2024 Fund\n\nInvestment…
3      American Funds College 2033 Fund\n\nInvestment…
4      American Funds College 2036 Fund\n\nInvestment…
..                                                   …
254    Fund Summary\nInvestment Objective\nThe fund s…
255    Fund Summary\nInvestment Objective\nThe fund s…
256    Fund Summary\nInvestment Objective\nThe fund s…
257    Fund Summary\nInvestment Objective\nThe fund s…
258    Fund Summary\nInvestment Objective\nThe fund s…

[259 rows x 3 columns]
```

```python
# Takes a summary, the knowledge base and some hyper parameters and returns the
→"num_sent" sentences
# of the summary that are closer to the the knowledge base in term of spacial
→distances.
def extract_sentence_distance(summary, knowledge, n_closer, n_reject, num_sent):
    # Split the summary into sentences.
    sentences = sent_tokenize(summary)
    sentence_scores = []
    # Loop over the sentences.
    for j, sentence in enumerate(sentences):
        # we tokenize and clean the sentence
        tokens = tokenizer(sentence)

        sentence_barycentre = np.zeros(embedding_size)
        effective_len = 0
        # Compute the barycentre of the sentence
        for token in tokens :
            try :
                sentence_barycentre += np.array(word2vec[token])
                effective_len += 1
            except KeyError :
                pass
            except :
                raise

        # Reject sentences with less than n_reject words in our word2vec map
        if effective_len <= n_reject :
            sentence_scores.append(1)

        else :
            sentence_barycentre = sentence_barycentre/effective_len
            # Compute the distance sentece_barycentre -> words in our knowledge
    →base
            barycentre_distance = [cosine(sentence_barycentre,
    →word2vec[key_word]) for key_word in knowledge]
```

```
            barycentre_distance.sort()
            # Create the score of the sentence by averaging the "n_closer"
    ↪smallest distances
            score = np.mean(barycentre_distance[:n_closer])
            sentence_scores.append(score)
    # Select the "num_sent" sentences that have the smallest score (smallest
    ↪distance score with the knowledge base)
    sentence_scores, sentences = zip(*sorted(zip(sentence_scores, sentences)))
    top_sentences = sentences[:num_sent]
    return ' '.join(top_sentences)
```

[81]:
```
# It takes several minutes (5-10 minutes)
df_main_train['strategy_sentences_distance'] = df_main_train.apply(lambda x :
 ↪extract_sentence_distance(x['summary'], knowledge_base, n_closer=10,
 ↪n_reject=5, num_sent=5), axis=1)
```

```
    ---------------------------------------------------------------------------
    KeyboardInterrupt                         Traceback (most recent call last)
    <ipython-input-81-653d2dc42d86> in <cell line: 0>()
          1 # It takes several minutes (5-10 minutes)
    ----> 2 df_main_train['strategy_sentences_distance'] = df_main_train.
     ↪apply(lambda x : extract_sentence_distance(x['summary'], knowledge_base,
     ↪n_closer=10, n_reject=5, num_sent=5), axis=1)

    /usr/local/lib/python3.11/dist-packages/pandas/core/frame.py in apply(self,
     ↪func, axis, raw, result_type, args, by_row, engine, engine_kwargs, **kwargs)
      10372                kwargs=kwargs,
      10373            )
    > 10374            return op.apply().__finalize__(self, method="apply")
      10375
      10376    def map(

    /usr/local/lib/python3.11/dist-packages/pandas/core/apply.py in apply(self)
        914                return self.apply_raw(engine=self.engine, engine_kwargs=self.
     ↪engine_kwargs)
        915
    --> 916            return self.apply_standard()
        917
        918    def agg(self):

    /usr/local/lib/python3.11/dist-packages/pandas/core/apply.py in
     ↪apply_standard(self)
       1061    def apply_standard(self):
       1062        if self.engine == "python":
    -> 1063            results, res_index = self.apply_series_generator()
       1064        else:
       1065            results, res_index = self.apply_series_numba()
```

```
/usr/local/lib/python3.11/dist-packages/pandas/core/apply.py in
 ↪apply_series_generator(self)
   1079                for i, v in enumerate(series_gen):
   1080                    # ignore SettingWithCopy here in case the user mutates
-> 1081                    results[i] = self.func(v, *self.args, **self.kwargs)
   1082                    if isinstance(results[i], ABCSeries):
   1083                        # If we have a view on v, we need to make a copy
 ↪because


<ipython-input-81-653d2dc42d86> in <lambda>(x)
      1 # It takes several minutes (5-10 minutes)
----> 2 df_main_train['strategy_sentences_distance'] = df_main_train.
 ↪apply(lambda x : extract_sentence_distance(x['summary'], knowledge_base,
 ↪n_closer=10, n_reject=5, num_sent=5), axis=1)


<ipython-input-76-84cc2b415a80> in extract_sentence_distance(summary, knowledge
 ↪n_closer, n_reject, num_sent)
     29                sentence_barycentre = sentence_barycentre/effective_len
     30                # Compute the distance sentece_barycentre -> words in our
 ↪knowledge base
---> 31                barycentre_distance = [cosine(sentence_barycentre,
 ↪word2vec[key_word]) for key_word in knowledge]
     32                barycentre_distance.sort()
     33                # Create the score of the sentence by averaging the
 ↪"n_closer" smallest distances


<ipython-input-76-84cc2b415a80> in <listcomp>(.0)
     29                sentence_barycentre = sentence_barycentre/effective_len
     30                # Compute the distance sentece_barycentre -> words in our
 ↪knowledge base
---> 31                barycentre_distance = [cosine(sentence_barycentre,
 ↪word2vec[key_word]) for key_word in knowledge]
     32                barycentre_distance.sort()
     33                # Create the score of the sentence by averaging the
 ↪"n_closer" smallest distances


/usr/local/lib/python3.11/dist-packages/scipy/spatial/distance.py in cosine(u,
 ↪v, w)
    736     # cosine distance is also referred to as 'uncentered correlation',
    737     #   or 'reflective correlation'
--> 738     return correlation(u, v, w=w, centered=False)
    739
    740


/usr/local/lib/python3.11/dist-packages/scipy/spatial/distance.py in
 ↪correlation(u, v, w, centered)
    685        dist = 1.0 - uv / math.sqrt(uu * vv)
```

```
     686         # Clip the result to avoid rounding error
--> 687         return np.clip(dist, 0.0, 2.0)
     688
     689

/usr/local/lib/python3.11/dist-packages/numpy/_core/fromnumeric.py in clip(a,
 ↪a_min, a_max, out, **kwargs)
    2245
    2246        """
 -> 2247        return _wrapfunc(a, 'clip', a_min, a_max, out=out, **kwargs)
    2248
    2249

/usr/local/lib/python3.11/dist-packages/numpy/_core/fromnumeric.py in
 ↪_wrapfunc(obj, method, *args, **kwds)
      55
      56        try:
---> 57            return bound(*args, **kwds)
      58        except TypeError:
      59            # A TypeError occurs if the object does have such a method in its

/usr/local/lib/python3.11/dist-packages/numpy/_core/_methods.py in _clip(a, min
 ↪max, out, **kwargs)
      97        return items
      98
---> 99 def _clip(a, min=None, max=None, out=None, **kwargs):
     100        if min is None and max is None:
     101            raise ValueError("One of max or min must be given")

KeyboardInterrupt:
```

```
[82]: df_main_train
```

```
[82]:                                       fund_name  Prediction  \
      0                American Funds College 2018 Fund          2.0
      1                American Funds College 2021 Fund          2.0
      2                American Funds College 2024 Fund          2.0
      3                American Funds College 2033 Fund          2.0
      4                American Funds College 2036 Fund          2.0
      ..                                            ...          ...
      254                  American Century Growth ETF          0.0
      255                   American Century Value ETF          0.0
      256  American Century Diversified Corporate Bond ETF          1.0
      257  American Century Diversified Municipal Bond ETF          1.0
      258    American Century STOXX U.S. Quality Value ETF          0.0

                                        summary  \
```

```
0      American Funds College 2018 Fund\n\nInvestment…
1      American Funds College 2021 Fund\n\nInvestment…
2      American Funds College 2024 Fund\n\nInvestment…
3      American Funds College 2033 Fund\n\nInvestment…
4      American Funds College 2036 Fund\n\nInvestment…
..                                                   …
254    Fund Summary\nInvestment Objective\nThe fund s…
255    Fund Summary\nInvestment Objective\nThe fund s…
256    Fund Summary\nInvestment Objective\nThe fund s…
257    Fund Summary\nInvestment Objective\nThe fund s…
258    Fund Summary\nInvestment Objective\nThe fund s…

                             strategy_sentences_distance
0      Equity-income and balanced funds generally str…
1      Equity-income and balanced funds generally str…
2      Equity-income and balanced funds generally str…
3      Equity-income and balanced funds generally str…
4      Equity-income and balanced funds generally str…
..                                                    …
254    The fund will generally issue and redeem Creat…
255    To the extent that a large shareholder (includ…
256    •\nCash Transactions Risk – The fund may effec…
257    •\nCash Transactions Risk – The fund may effec…
258    An income sustainability screen based on divid…

[259 rows x 4 columns]
```

```python
[83]: df_main_val['strategy_sentences_distance'] = df_main_val.apply(lambda x :
      ↪extract_sentence_distance(x['summary'], knowledge_base, n_closer=10,
      ↪n_reject=5, num_sent=5), axis=1)
```

```python
[84]: df_main_test['strategy_sentences_distance'] = df_main_test.apply(lambda x :
      ↪extract_sentence_distance(x['summary'], knowledge_base, n_closer=10,
      ↪n_reject=5, num_sent=5), axis=1)
```

```python
[85]: df_main_val
```

```
[85]:                                       fund_name  Prediction  \
0                American Funds College 2027 Fund         2.0
1                American Funds College 2030 Fund         2.0
2           American Funds College Enrollment Fund       1.0
3     American Funds U.S. Government Money Market Fund  1.0
4                            Global Balanced Fund        2.0
..                                            …           …
81           Anchor Tactical Credit Strategies Fund     1.0
82              Federated MDT Small Cap Growth Fund     0.0
83                            S&P 500 Index Fund        0.0
```

```
84                          Managed Allocation Fund        2.0
85   American Century STOXX U.S. Quality Growth ETF        0.0

                                     summary  \
0    American Funds College 2027 Fund\n\nInvestment…
1    American Funds College 2030 Fund\n\nInvestment…
2    American Funds College Enrollment Fund\n\nInve…
3    Investment objective The investment objective …
4    Global Balanced Fund\n\nInvestment objectives …
..                                                …
81   FUND SUMMARY - ANCHOR TACTICAL CREDIT STRATEGI…
82   Fund Summary Information\nFederated MDT Small …
83   INVESTMENT OBJECTIVE\nThe USAA S&P 500 Index F…
84   INVESTMENT OBJECTIVE\nThe USAA Managed Allocat…
85   Fund Summary\nInvestment Objective\nThe fund s…

                           strategy_sentences_distance
0    Equity-income and balanced funds generally str…
1    Equity-income and balanced funds generally str…
2    Issuers of these securities may be more suscep…
3    Although your actual costs may be higher or lo…
4    Falling interest rates may cause an issuer to …
..                                                …
81   A general downturn in the securities market ma…
82   The Adviser implements its strategy using a qu…
83   Although your actual costs may be higher or lo…
84   In addition, market developments and other fac…
85   •\nSampling   Risk - To the extent the fund us…

[86 rows x 4 columns]
```

[90]: `df_main_test`

```
[90]:                                     fund_name  Prediction  \
0                                   Bond Fund        1.0
1                         Corporate Bond Fund        1.0
2                            Global Bond Fund        1.0
3                 Global Growth and Income Fund        2.0
4              Global Small Capitalization Fund        0.0
..                                          …          …
82               Federated MDT All Cap Core Fund        0.0
83                         Nasdaq-100 Index Fund        0.0
84                            Income Stock Fund        0.0
85   USAA MSCI International Value Momentum Blend I…        0.0
86   American Century Quality Diversified Internati…        0.0

                                     summary  \
```

```
0    Bond Fund\n\nInvestment objective The fund's i…
1    Investment objective The fund's investment obj…
2    Global Bond Fund\n\nInvestment objective The f…
3    Global Growth and Income Fund\n\nInvestment ob…
4    Global Small Capitalization Fund\n\nInvestment…
..                                                 …
82   Fund Summary Information\nFederated MDT All Ca…
83   INVESTMENT OBJECTIVE\nThe USAA Nasdaq-100 Inde…
84   INVESTMENT OBJECTIVE\nThe USAA Income Stock Fu…
85   INVESTMENT OBJECTIVE\nThe USAA MSCI USA Value …
86   Fund Summary\nInvestment Objective\nThe fund s…

                           strategy_sentences_distance
0    Falling interest rates may cause an issuer to …
1    Falling interest rates may cause an issuer to …
2    Although the fund does not intend to limit its…
3    Issuers of these securities may be more suscep…
4    Issuers of these securities may be more suscep…
..                                                 …
82   3\nTable of Contents\n(For the Period Ended De…
83   AVERAGE ANNUAL TOTAL RETURNS\nFor Periods Ende…
84   Although your actual costs may be higher or lo…
85   These securities may be subject to more abrupt…
86   The market value of depositary receipts is dep…

[87 rows x 4 columns]
```

```python
#def extract_sentence_match(summary, knowledge, num_sent):
#    sentences = sent_tokenize(summary)
#    sentence_scores = []
#    for j, sentence in enumerate(sentences):
#        set_tokens = set(tokenizer(sentence))
#
#        # Find the number of common words between the knowledge base and the
 ↪sentence
 #        inter_knwoledge = set_tokens.intersection(knowledge)
#
 #        sentence_scores.append(len(inter_knwoledge))
#
 #   sentence_scores, sentences = zip(*sorted(zip(sentence_scores, sentences)))
  #   top_sentences = sentences[len(sentences)-num_sent-1:]
   # return ' '.join(top_sentences)
```

```python
#df_main['strategy_sentences_match'] = df_main.apply(lambda x :
 ↪extract_sentence_match(x['Summary'], knowledge_base, num_sent=5), axis=1)
```

```python
df_main_train
```

```
[86]:                                          fund_name   Prediction  \
      0              American Funds College 2018 Fund          2.0
      1              American Funds College 2021 Fund          2.0
      2              American Funds College 2024 Fund          2.0
      3              American Funds College 2033 Fund          2.0
      4              American Funds College 2036 Fund          2.0
      ..                                         …            …
      254                   American Century Growth ETF         0.0
      255                    American Century Value ETF         0.0
      256  American Century Diversified Corporate Bond ETF    1.0
      257  American Century Diversified Municipal Bond ETF    1.0
      258     American Century STOXX U.S. Quality Value ETF   0.0

                                               summary  \
      0    American Funds College 2018 Fund\n\nInvestment…
      1    American Funds College 2021 Fund\n\nInvestment…
      2    American Funds College 2024 Fund\n\nInvestment…
      3    American Funds College 2033 Fund\n\nInvestment…
      4    American Funds College 2036 Fund\n\nInvestment…
      ..                                             …
      254  Fund Summary\nInvestment Objective\nThe fund s…
      255  Fund Summary\nInvestment Objective\nThe fund s…
      256  Fund Summary\nInvestment Objective\nThe fund s…
      257  Fund Summary\nInvestment Objective\nThe fund s…
      258  Fund Summary\nInvestment Objective\nThe fund s…

                              strategy_sentences_distance
      0    Equity-income and balanced funds generally str…
      1    Equity-income and balanced funds generally str…
      2    Equity-income and balanced funds generally str…
      3    Equity-income and balanced funds generally str…
      4    Equity-income and balanced funds generally str…
      ..                                             …
      254  The fund will generally issue and redeem Creat…
      255  To the extent that a large shareholder (includ…
      256  •\nCash Transactions Risk – The fund may effec…
      257  •\nCash Transactions Risk – The fund may effec…
      258  An income sustainability screen based on divid…

      [259 rows x 4 columns]

[ ]:
```

0.5 Information extraction using deep classifiers

```
[79]: #X = df_main['strategy_sentences_match'].values
      X_train = df_main_train['strategy_sentences_distance'].values # uncomment to
       ↪use the first sentence extraction method.
      # Clean the texts
      X_train = [' '.join(tokenizer(txt)) for txt in X_train]
      X_train[0][:100]
```

```
[79]: 'equity-income balanced funds generally strive income growth stocks and/or bond
      investments bond fund'
```

```
[87]: #X = df_main['strategy_sentences_match'].values
      X_val = df_main_val['strategy_sentences_distance'].values # uncomment to use
       ↪the first sentence extraction method.
      # Clean the texts
      X_val = [' '.join(tokenizer(txt)) for txt in X_val]
      X_val[0][:100]
```

```
[87]: 'equity-income balanced funds generally strive income growth stocks and/or bond
      investments bond fund'
```

```
[88]: #X = df_main['strategy_sentences_match'].values
      X_test = df_main_test['strategy_sentences_distance'].values # uncomment to use
       ↪the first sentence extraction method.
      # Clean the texts
      X_test = [' '.join(tokenizer(txt)) for txt in X_test]
      X_test[0][:100]
```

```
[88]: 'falling interest rates may cause issuer redeem call refinance debt security
      stated maturity may resu'
```

```
[91]: y_train = df_main_train['Prediction']
      y_train
```

```
[91]: 0      2.0
      1      2.0
      2      2.0
      3      2.0
      4      2.0
             …
      254    0.0
      255    0.0
      256    1.0
      257    1.0
      258    0.0
      Name: Prediction, Length: 259, dtype: float64
```

```
[92]: y_val = df_main_val['Prediction']
      y_val
```

```
[92]: 0     2.0
      1     2.0
      2     1.0
      3     1.0
      4     2.0
           ...
      81    1.0
      82    0.0
      83    0.0
      84    2.0
      85    0.0
      Name: Prediction, Length: 86, dtype: float64
```

```
[93]: y_test = df_main_test['Prediction']
      y_test
```

```
[93]: 0     1.0
      1     1.0
      2     1.0
      3     2.0
      4     0.0
           ...
      82    0.0
      83    0.0
      84    0.0
      85    0.0
      86    0.0
      Name: Prediction, Length: 87, dtype: float64
```

0.6 Data Pre-Processing

```
[94]: # In order to set the length of the document series (maxlen), we visualise the␣
      ↪distribution of the number of words per documents.
      sizes = []
      out = display(progress(0, len(X_train)), display_id=True)
      for ii, txt in enumerate(X_train):
          sizes.append(len(word_tokenize(txt)))
          out.update(progress(ii, len(X_train)))

      sns.distplot(sizes, kde=False, rug=True)
```

```
<IPython.core.display.HTML object>
```

```
<ipython-input-94-bc141d98f3b4>:8: UserWarning:
```

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

```
  sns.distplot(sizes, kde=False, rug=True)
```

[94]: <Axes: >
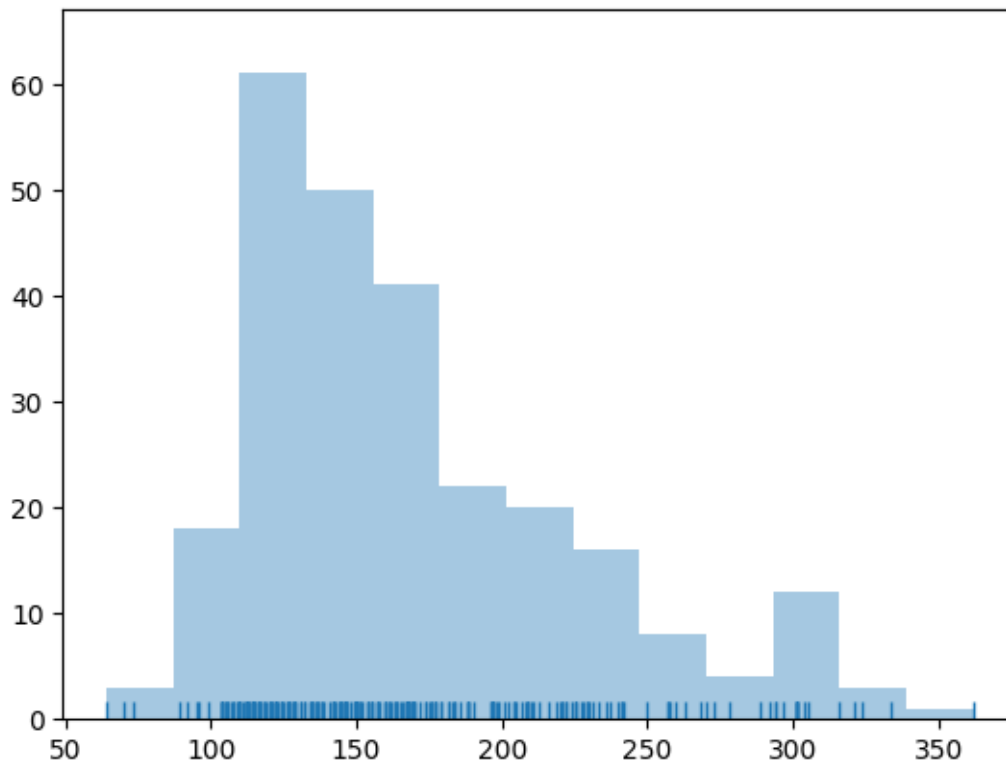

```
[140]: num_words = 2500 # Size of the vocabulary used. we only consider the 2500 most␣
        ↪common words. The other words are removed from the texts.
       maxlen = 200 # Number of word considered for each document. we cut or lengthen␣
        ↪the texts to have texts of 150 words.
       word_dimension = 50 # dimension of our word vectors.
```

```
[141]: # The Tokenizer provided by the Keras library allows to perform such␣
        ↪transformation.
       keras_tokenizer = Tokenizer(num_words=num_words)
       keras_tokenizer.fit_on_texts(X_train)
```

```
# word_index is the dictionary that contains the index of each words in our␣
  ↪2500 long vocabulary.
word_index = keras_tokenizer.word_index
sequences_train = keras_tokenizer.texts_to_sequences(X_train)
sequences_val = keras_tokenizer.texts_to_sequences(X_val)
sequences_test=keras_tokenizer.texts_to_sequences(X_test)
```

[142]:
```
print('Text : '+X_train[0])
print('Index sequences : ' +str(sequences_train[0]))
```

Text : equity-income balanced funds generally strive income growth stocks and/or
bond investments bond funds seek current income bond investments securities
rated bb+ ba1 sometimes referred junk bonds. exposure lower rated securities may
help fund achieve objective providing current income falling interest rates may
cause issuer redeem call refinance debt security stated maturity may result fund
reinvest proceeds lower yielding securities issuers securities may susceptible
actions foreign governments imposition price controls punitive taxes could
adversely impact value securities selection underlying funds allocation fund
assets could cause fund lose value results lag relevant benchmarks funds similar
objectives
Index sequences : [50, 29, 457, 16, 162, 687, 29, 84, 111, 39, 40, 49, 34, 49,
16, 458, 213, 29, 49, 34, 9, 261, 607, 608, 527, 484, 567, 185, 87, 68, 261, 9,
3, 328, 1, 322, 251, 568, 213, 29, 323, 46, 57, 3, 95, 206, 90, 304, 387, 171,
54, 252, 313, 3, 88, 1, 342, 207, 68, 324, 9, 58, 9, 3, 243, 334, 82, 228, 296,
38, 266, 343, 2, 31, 103, 121, 15, 9, 237, 104, 16, 174, 1, 65, 31, 95, 1, 224,
15, 96, 283, 194, 255, 16, 214, 229]

[143]:
```
len(sequences_train)
```

[143]: 259

[144]:
```
#Then we use the pad_sequence function provided by keras to truncate or␣
  ↪lengthen (by adding zeros)
#each text so they all have the same length : 170
feature_train = pad_sequences(sequences_train, maxlen=maxlen, dtype=float,␣
  ↪padding='post', truncating='post')
feature_val = pad_sequences(sequences_val, maxlen=maxlen, dtype=float,␣
  ↪padding='post', truncating='post')
feature_test=pad_sequences(sequences_test, maxlen=maxlen, dtype=float,␣
  ↪padding='post', truncating='post')
```

[145]:
```
print(len(X_train[0]))
len(feature_train[0])
```

714

[145]: 200

```
[146]: word2vec.get("stock")
```

```
[146]: array([-0.52702427,  0.546883  , -0.5217987 , -0.43095547, -0.5527872 ,
               0.4939101 ,  0.456268  , -0.50855875, -0.5313279 , -0.48267663,
              -0.49233153,  0.47901547,  0.48313826, -0.4651238 , -0.46131042,
              -0.4708229 ,  0.498532  ,  0.4933306 , -0.5404466 ,  0.47832614,
              -0.45568234,  0.46188068,  0.44874865, -0.46084365, -0.43395746,
               0.5326104 ,  0.45170507, -0.49335197,  0.5231367 ,  0.43798825,
               0.4742455 , -0.47718424, -0.4704817 , -0.50724834, -0.43765053,
              -0.4610433 ,  0.49027053, -0.49304688, -0.4748721 , -0.4476859 ,
               0.49050564,  0.49190056,  0.51036674, -0.4561873 ,  0.4454228 ,
               0.5108056 , -0.47885454, -0.5172404 , -0.44386956,  0.48884946],
             dtype=float32)
```

```
[147]: word_index.get("investments")
```

```
[147]: 34
```

```
[148]: X_train[0]
```

```
[148]: 'equity-income balanced funds generally strive income growth stocks and/or bond
       investments bond funds seek current income bond investments securities rated bb+
       ba1 sometimes referred junk bonds. exposure lower rated securities may help fund
       achieve objective providing current income falling interest rates may cause
       issuer redeem call refinance debt security stated maturity may result fund
       reinvest proceeds lower yielding securities issuers securities may susceptible
       actions foreign governments imposition price controls punitive taxes could
       adversely impact value securities selection underlying funds allocation fund
       assets could cause fund lose value results lag relevant benchmarks funds similar
       objectives'
```

```
[149]: print ('length of the first sequence : '+str(len(sequences_train[0])) )
       print ('length of the paded first sequence : '+str(len(feature_train[0])) )

       length of the first sequence : 96
       length of the paded first sequence : 200
```

```
[150]: #matrix representation of the words in our vocabulary
       # word_index is the dictionary that contains the index of each words in our
       ↪2500 long vocabulary.
       embedding_matrix = np.zeros((len(word_index) + 1, word_dimension))
       for word, i in word_index.items():
           embedding_vector = word2vec.get(word)

           if embedding_vector is not None:
               embedding_matrix[i] = embedding_vector
```

```
[151]: embedding_matrix[1]
```

```
[151]: array([-0.51161611,  0.49861488, -0.45986888, -0.51044023, -0.4733614 ,
                0.48179847,  0.48942891, -0.52708107, -0.47881442, -0.44727966,
               -0.45656052,  0.4437072 ,  0.45148158, -0.47779301, -0.51318449,
               -0.47875041,  0.52369714,  0.45715469, -0.49631903,  0.52710831,
               -0.48395544,  0.5100708 ,  0.50592208, -0.46686277, -0.50518817,
                0.509      ,  0.43394583, -0.44495079,  0.45079893,  0.49803144,
                0.51884866, -0.4787567 , -0.46801475, -0.52083337, -0.500175  ,
               -0.53125322,  0.42735058, -0.47766268, -0.46364808, -0.49993426,
                0.46382937,  0.51606905,  0.50196582, -0.50648367,  0.51581132,
                0.44917157, -0.4634071 , -0.51167554, -0.50772369,  0.48334512])
```

```
[152]: embedding_matrix.shape
```

```
[152]: (2201, 50)
```

```
[153]: len(word_index) + 1
```

```
[153]: 2201
```

```
[154]: word_dimension
```

```
[154]: 50
```

0.7   First model: One dimensional Convolutional Neural Network

def create_CNN_model(): CNN = Sequential() # The Embedding layer takes the embedding matrix as an argument and transform the inputed the sequences of index to sequences of vectors. CNN.add(Embedding(len(word_index) + 1, word_dimension, weights=[embedding_matrix], input_length=maxlen, trainable=False))

```
CNN.add(Convolution1D(64, 5, activation = 'relu'))
CNN.add(MaxPooling1D(pool_size = 5))

CNN.add(Convolution1D(32, 5, activation = 'relu'))
CNN.add(MaxPooling1D(pool_size = 5))

CNN.add(Flatten())
CNN.add(Dense(units = 128 , activation = 'relu'))
CNN.add(Dropout(0.5))
CNN.add(Dense(units = 1, activation = 'sigmoid'))

CNN.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])

return CNN
```

```
[155]:  def create_CNN_model():
            CNN = Sequential()
            CNN.add(Embedding(len(word_index) + 1, word_dimension,␣
        ↪weights=[embedding_matrix], input_length=maxlen, trainable=False))

            CNN.add(Convolution1D(64, 5, activation='relu'))
            CNN.add(MaxPooling1D(pool_size=5))

            CNN.add(Convolution1D(32, 5, activation='relu'))
            CNN.add(MaxPooling1D(pool_size=5))

            CNN.add(Flatten())
            CNN.add(Dense(units=128, activation='relu'))
            CNN.add(Dropout(0.5))
            CNN.add(Dense(units=3, activation='softmax'))

            CNN.compile(optimizer='adam', loss='sparse_categorical_crossentropy',␣
        ↪metrics=['accuracy'])

            return CNN
```

```
[156]:  CNN_model = create_CNN_model()
```

/usr/local/lib/python3.11/dist-packages/keras/src/layers/core/embedding.py:90:
UserWarning: Argument `input_length` is deprecated. Just remove it.
  warnings.warn(

```
[157]:  CNN_history = CNN_model.fit(feature_train, y_train, epochs=10, batch_size=16,␣
        ↪validation_data=(feature_val, y_val))
```

```
Epoch 1/10
17/17                6s 182ms/step -
accuracy: 0.4377 - loss: 1.1138 - val_accuracy: 0.5116 - val_loss: 1.0357
Epoch 2/10
17/17                1s 6ms/step -
accuracy: 0.4808 - loss: 1.0523 - val_accuracy: 0.5116 - val_loss: 1.0059
Epoch 3/10
17/17                0s 8ms/step -
accuracy: 0.5401 - loss: 0.9956 - val_accuracy: 0.5349 - val_loss: 0.9953
Epoch 4/10
17/17                0s 6ms/step -
accuracy: 0.4963 - loss: 1.0032 - val_accuracy: 0.5233 - val_loss: 0.9575
Epoch 5/10
17/17                0s 6ms/step -
accuracy: 0.5382 - loss: 0.9328 - val_accuracy: 0.5116 - val_loss: 0.9200
Epoch 6/10
17/17                0s 6ms/step -
```

```
accuracy: 0.5751 - loss: 0.9207 - val_accuracy: 0.5349 - val_loss: 0.9115
Epoch 7/10
17/17                0s 6ms/step -
accuracy: 0.5232 - loss: 0.9376 - val_accuracy: 0.5581 - val_loss: 0.9057
Epoch 8/10
17/17                0s 8ms/step -
accuracy: 0.6285 - loss: 0.8496 - val_accuracy: 0.5581 - val_loss: 0.8928
Epoch 9/10
17/17                0s 6ms/step -
accuracy: 0.6083 - loss: 0.8763 - val_accuracy: 0.5581 - val_loss: 0.8897
Epoch 10/10
17/17                0s 6ms/step -
accuracy: 0.5759 - loss: 0.9007 - val_accuracy: 0.5233 - val_loss: 0.9294
```

[158]:
```python
# Plot training & validation accuracy
plt.plot(CNN_history.history['accuracy'])
plt.plot(CNN_history.history['val_accuracy'])
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Train', 'Validation'], loc='upper left')
plt.show()
```
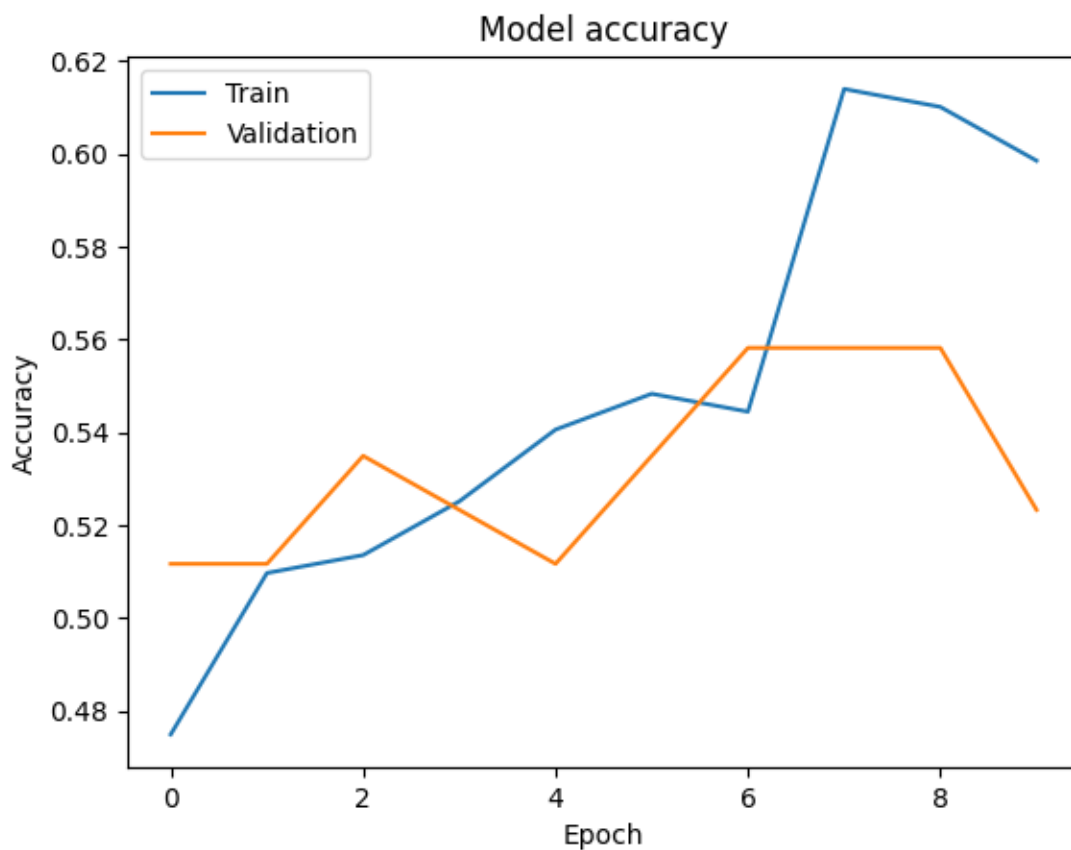
```
[159]:  # Plot training & validation loss values
        plt.plot(CNN_history.history['loss'])
        plt.plot(CNN_history.history['val_loss'])
        plt.title('Model loss')
        plt.ylabel('Loss')
        plt.xlabel('Epoch')
        plt.legend(['Train', 'Validation'], loc='upper left')
        plt.show()
```
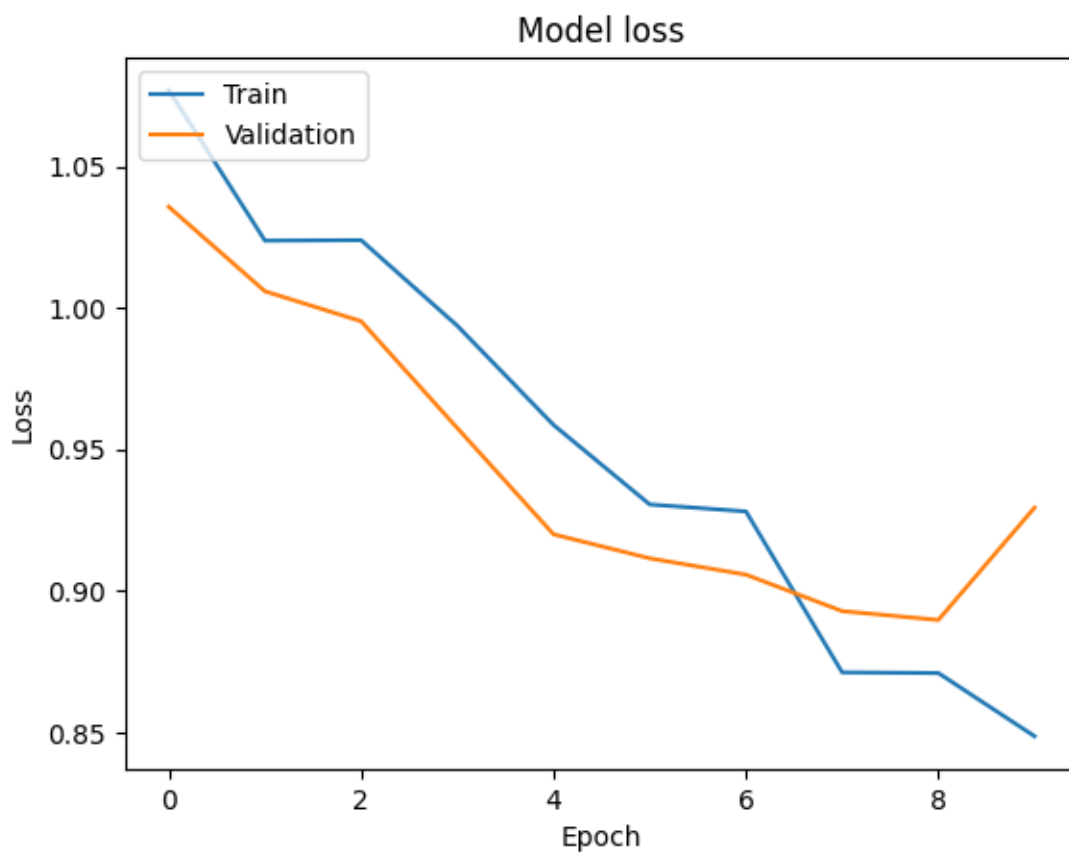


```
[161]:  y_pred_CNN = CNN_model.predict(feature_val)
```

        3/3   0s 8ms/step

```
[162]:  y_pred_label = np.argmax(y_pred_CNN, axis=1)
```

```
[163]:  report=classification_report(y_val,y_pred_label)
        print(report)
```

	precision	recall	f1-score	support
0.0	0.53	0.95	0.68	44
1.0	0.00	0.00	0.00	23
2.0	0.50	0.16	0.24	19
accuracy			0.52	86
macro avg	0.34	0.37	0.31	86
weighted avg	0.38	0.52	0.40	86

```
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

0.8 Second Model

```python
[121]: def create_RNN_model():
           RNN = Sequential()
           RNN.add(Embedding(len(word_index) + 1, word_dimension,
       ↪weights=[embedding_matrix], input_length = maxlen, trainable=False))

           RNN.add(Bidirectional(LSTM(word_dimension)))
           RNN.add(Dense(word_dimension, activation='relu'))
           RNN.add(Dense(1, activation='sigmoid'))
           RNN.compile(loss='binary_crossentropy', optimizer='Adam',
       ↪metrics=['accuracy'])

           return RNN
```

```python
[122]: RNN_model = create_RNN_model()
```

```
/usr/local/lib/python3.11/dist-packages/keras/src/layers/core/embedding.py:90:
UserWarning: Argument `input_length` is deprecated. Just remove it.
  warnings.warn(
```

```
[123]: RNN_history = RNN_model.fit(feature_train, y_train, epochs=10, batch_size=16,
        ↪validation_data=(feature_val, y_val))
```

```
Epoch 1/10
17/17              5s 41ms/step -
accuracy: 0.2814 - loss: 0.6618 - val_accuracy: 0.2674 - val_loss: 0.5685
Epoch 2/10
17/17              0s 29ms/step -
accuracy: 0.2620 - loss: 0.6315 - val_accuracy: 0.2674 - val_loss: 0.5798
Epoch 3/10
17/17              1s 25ms/step -
accuracy: 0.2785 - loss: 0.6273 - val_accuracy: 0.2674 - val_loss: 0.5816
Epoch 4/10
17/17              1s 31ms/step -
accuracy: 0.2715 - loss: 0.6117 - val_accuracy: 0.2674 - val_loss: 0.5654
Epoch 5/10
17/17              0s 24ms/step -
accuracy: 0.2859 - loss: 0.6053 - val_accuracy: 0.2674 - val_loss: 0.5819
Epoch 6/10
17/17              1s 27ms/step -
accuracy: 0.2736 - loss: 0.5916 - val_accuracy: 0.2674 - val_loss: 0.5642
Epoch 7/10
17/17              0s 18ms/step -
accuracy: 0.2882 - loss: 0.5757 - val_accuracy: 0.2674 - val_loss: 0.5690
Epoch 8/10
17/17              0s 23ms/step -
accuracy: 0.2676 - loss: 0.6535 - val_accuracy: 0.2674 - val_loss: 0.5725
Epoch 9/10
17/17              0s 17ms/step -
accuracy: 0.2687 - loss: 0.5083 - val_accuracy: 0.2674 - val_loss: 0.5594
Epoch 10/10
17/17              0s 17ms/step -
accuracy: 0.3005 - loss: 0.6018 - val_accuracy: 0.2674 - val_loss: 0.5869
```

```
[124]: # Plot training & validation accuracy values
       plt.plot(RNN_history.history['accuracy'])
       plt.plot(RNN_history.history['val_accuracy'])
       plt.title('Model accuracy')
       plt.ylabel('Accuracy')
       plt.xlabel('Epoch')
       plt.legend(['Train', 'Validation'], loc='upper left')
       plt.show()
```

Model accuracy
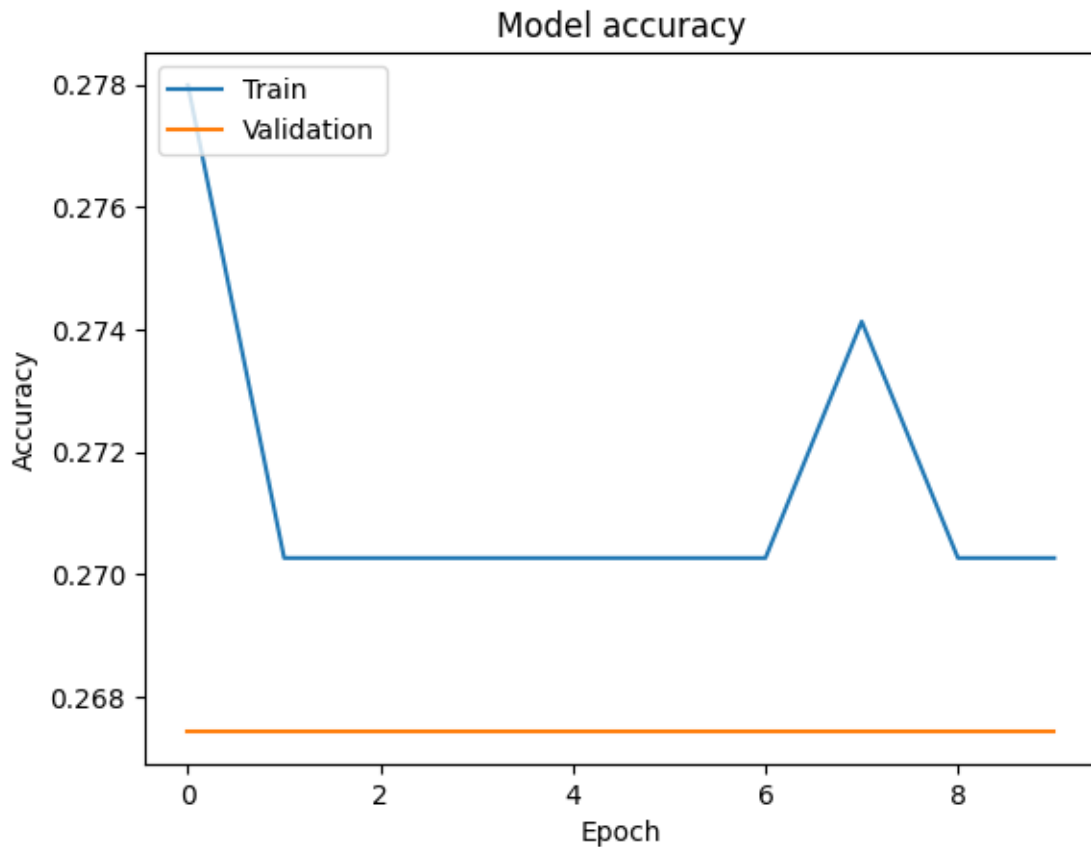
0.9 New word2vec

```
word2vec_new=load_word2vec("/content/gdrive/MyDrive/Colab Notebooks/MF815/NLP/
↪NLP_app/word2vec_perso.txt")
```

```
#matrix representation of the words in our vocabulary
# word_index is the dictionary that contains the index of each words in our␣
↪2500 long vocabulary.
embedding_matrix_new = np.zeros((len(word_index) + 1, word_dimension))
for word, i in word_index.items():
    embedding_vector = word2vec_new.get(word)

    if embedding_vector is not None:
        embedding_matrix_new[i] = embedding_vector
```

0.10 Transform document txt to vector

```
[164]: def document_vector_mean(feature_data, embedding_matrix):
           doc_vectors = []
           for doc in feature_data:
               word_vecs = []
               for idx in doc:
                   if idx != 0:
                       word_vec = embedding_matrix[idx]
                       word_vecs.append(word_vec)
               if word_vecs:
                   doc_vector = np.mean(word_vecs, axis=0)
               else:
                   doc_vector = np.zeros(embedding_matrix.shape[1])
               doc_vectors.append(doc_vector)
           return np.array(doc_vectors)
```

```
[165]: def document_vector_norm_weighted(feature_data, embedding_matrix):
           doc_vectors = []
           for doc in feature_data:
               word_vecs = [embedding_matrix[idx] for idx in doc if idx != 0]
               if word_vecs:
                   word_arr = np.array(word_vecs)
                   norms = np.linalg.norm(word_arr, axis=1)
                   sum_norms = norms.sum()
                   if sum_norms == 0:
                       doc_vector = np.mean(word_arr, axis=0)  # Fallback to mean
                   else:
                       weights = norms / sum_norms
                       doc_vector = np.dot(weights, word_arr)
               else:
                   doc_vector = np.zeros(embedding_matrix.shape[1])
               doc_vectors.append(doc_vector)
           return np.array(doc_vectors)
```

```
[166]: feature_train_int = feature_train.astype(int)
       feature_val_int=feature_val.astype(int)
       feature_test_int=feature_test.astype(int)
       X_train_vec=document_vector_mean(feature_train_int, embedding_matrix)
       X_val_vec=document_vector_mean(feature_val_int, embedding_matrix)
       X_test_vec=document_vector_mean(feature_test_int, embedding_matrix)
```

```
[167]: X_train_vec=document_vector_norm_weighted(feature_train_int, embedding_matrix)
       X_val_vec=document_vector_norm_weighted(feature_val_int, embedding_matrix)
       X_test_vec=document_vector_norm_weighted(feature_test_int, embedding_matrix)
```

```
[ ]: X_train_vec=document_vector_norm_weighted(feature_train_int,␣
     ↪embedding_matrix_new)
     X_val_vec=document_vector_norm_weighted(feature_val_int, embedding_matrix_new)
     X_test_vec=document_vector_norm_weighted(feature_test_int, embedding_matrix_new)
```

```
[168]: X_train_vec
```

```
[168]: array([[-0.4801011 ,  0.49525288, -0.52563552, …, -0.4736114 ,
               -0.48246902,  0.47728673],
              [-0.47958325,  0.49537869, -0.52638409, …, -0.47432019,
               -0.48359371,  0.47665846],
              [-0.47958325,  0.49537869, -0.52638409, …, -0.47432019,
               -0.48359371,  0.47665846],
              …,
              [-0.48141966,  0.4890497 , -0.52176607, …, -0.46657929,
               -0.48199077,  0.47881029],
              [-0.47767958,  0.48654023, -0.52420494, …, -0.46318924,
               -0.4765423 ,  0.47752866],
              [-0.47847556,  0.4909724 , -0.5185461 , …, -0.47229304,
               -0.48389164,  0.47661523]])
```

0.11 SVM

```
[169]: from sklearn.model_selection import RandomizedSearchCV, StratifiedKFold
       from sklearn.svm import SVC
       from sklearn.metrics import accuracy_score, confusion_matrix,␣
       ↪classification_report
       from scipy.stats import uniform, reciprocal
```

```
[170]: from sklearn.svm import SVC
       liner_svm=SVC(kernel="linear",C=0.1,probability=True)
       liner_svm.fit(X_train_vec,y_train)
```

```
[170]: SVC(C=0.1, kernel='linear', probability=True)
```

```
[171]: svc_val_pred=liner_svm.predict(X_val_vec)
```

```
[172]: print("\nClassification Report:test set")
       print(classification_report(y_val, svc_val_pred))
```

```
Classification Report:test set
              precision    recall  f1-score   support

         0.0       0.51      1.00      0.68        44
         1.0       0.00      0.00      0.00        23
         2.0       0.00      0.00      0.00        19
```

```
    accuracy                                0.51        86
   macro avg         0.17        0.33       0.23        86
weighted avg         0.26        0.51       0.35        86
```

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples. Use `zero_division` parameter to control this
behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

```python
[173]: param_distributions = {
           'C': reciprocal(0.1, 100),
           'kernel': ['linear', 'rbf', 'poly'],
           'gamma': ['scale', 'auto', 0.001, 0.01, 0.1, 1],
           'degree': [2, 3, 4],
           'class_weight': [None, 'balanced'],
           'shrinking': [True, False]
       }

       svm = SVC()
       cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

       random_search = RandomizedSearchCV(
           svm, param_distributions=param_distributions,
           n_iter=50, cv=cv, scoring='accuracy', verbose=2, n_jobs=-1, random_state=42
       )
       random_search.fit(X_train_vec, y_train)
```

Fitting 5 folds for each of 50 candidates, totalling 250 fits

```
[173]: RandomizedSearchCV(cv=StratifiedKFold(n_splits=5, random_state=42,
       shuffle=True),
                          estimator=SVC(), n_iter=50, n_jobs=-1,
                          param_distributions={'C':
       <scipy.stats._distn_infrastructure.rv_continuous_frozen object at
       0x7a92b0588950>,
```

```
                              'class_weight': [None, 'balanced'],
                              'degree': [2, 3, 4],
                              'gamma': ['scale', 'auto', 0.001, 0.01,
                                      0.1, 1],
                              'kernel': ['linear', 'rbf', 'poly'],
                              'shrinking': [True, False]},
                  random_state=42, scoring='accuracy', verbose=2)
```

[174]:
```python
X_trainval_features = np.vstack([X_train_vec, X_val_vec])
y_trainval = np.hstack([y_train, y_val])


final_svm = SVC(**random_search.best_params_)
final_svm.fit(X_trainval_features, y_trainval)
```

[174]:
```
SVC(C=np.float64(65.84106160121607), class_weight='balanced', gamma=1,
    kernel='poly', shrinking=False)
```

[175]:
```python
y_test_pred = final_svm.predict(X_test_vec)
y_test_pred
```

[175]:
```
array([0., 0., 0., 2., 2., 2., 2., 0., 2., 2., 2., 2., 2., 2., 2., 2., 2.,
       1., 1., 1., 1., 1., 0., 0., 0., 1., 0., 1., 1., 1., 0., 1., 1., 1.,
       1., 1., 0., 0., 0., 0., 0., 0., 1., 1., 1., 1., 1., 1., 2., 0., 0.,
       1., 0., 0., 0., 0., 2., 1., 0., 0., 0., 1., 1., 1., 2., 2., 2., 0.,
       0., 0., 0., 1., 1., 2., 2., 1., 1., 1., 0., 0., 0., 0., 0., 2., 2.,
       0., 2.])
```

[176]:
```python
report = classification_report(y_test, y_test_pred)
print(report)
```

```
              precision    recall  f1-score   support

         0.0       0.76      0.70      0.73        37
         1.0       0.73      0.79      0.76        28
         2.0       0.74      0.77      0.76        22

    accuracy                           0.75        87
   macro avg       0.75      0.75      0.75        87
weighted avg       0.75      0.75      0.75        87
```

[177]:
```python
confusion_matrix(y_test, y_test_pred)
```

[177]:
```
array([[26,  6,  5],
       [ 5, 22,  1],
       [ 3,  2, 17]])
```