BPE_training1

↪ **input**

hyperparams.csv

output →

data/concatenated_UniProt_hp.txt

BPE_training2

↪ **input**

data/concatenated_UniProt_hp.txt

hyperparams.csv

output →

results/encoded_sequence/BPE_model_hp.bpe

generate_tokens

↪ **input**

hyperparams.csv

results/encoded_sequence/BPE_model_hp.bpe

output →

results/encoded_sequence/model_vocab_hp.csv

results/encoded_sequence/words_hp.csv

seq2vec_skipgrams

↪ **input**

hyperparams.csv

results/encoded_sequence/words_hp.csv

output →

results/embedded_sequence/seq2vec_ids.csv

results/embedded_sequence/skipgrams.txt

TF_IDF

↪ **input**

results/encoded_sequence/words_hp.csv

output →

results/encoded_sequence/TF_IDF_hp.csv

seq2vec_training

↪ **input**

hyperparams.csv

results/embedded_sequence/seq2vec_ids.csv

results/embedded_sequence/skipgrams.txt

output →

results/embedded_sequence/model.h5

results/embedded_sequence/model_metrics.txt

results/embedded_sequence/seq2vec_weights.csv

sequence_repres

↪ **input**

hyperparams.csv

results/embedded_sequence/model.h5

results/embedded_sequence/seq2vec_ids.csv

results/encoded_sequence/TF_IDF_hp.csv

results/encoded_sequence/words_hp.csv

output →

results/embedded_sequence/sequence_repres_seq2vec-SIF.csv

results/embedded_sequence/sequence_repres_seq2vec-TFIDF.csv

results/embedded_sequence/sequence_repres_seq2vec.csv

CCR

↪ **input**

results/embedded_sequence/sequence_repres_seq2vec-SIF.csv

results/embedded_sequence/sequence_repres_seq2vec-TFIDF.csv

results/embedded_sequence/sequence_repres_seq2vec.csv

output →

results/embedded_sequence/sequence_repres_seq2vec-SIF_CCR.csv

results/embedded_sequence/sequence_repres_seq2vec-TFIDF_CCR.csv

results/embedded_sequence/sequence_repres_seq2vec_CCR.csv

all

↪ **input**

results/embedded_sequence/model.h5

results/embedded_sequence/sequence_repres_seq2vec.csv

results/embedded_sequence/sequence_repres_seq2vec_CCR.csv