The Super Duper NLP Repo

Quantum stat

updated 12.02.20 ⚡ 314 notebooks

search...

1 2 3 4

Notebook	Added	Description	Model	Task	Creator	Source
Noun Phrase Chunker Adapter	12.02.20	Demo for building an adapter for conducting inference on chunk phrase extraction task	BERT Adapter	Chunk Phrase Extraction	AdapterHub	CO Open in Colab
Model Sparsification	12.02.20	Demo for sparsifying a RoBERTa model prior to training in PyTorch	RoBERTa	Sparsification	Hugging Face	CO Open in Colab
LXMERT	12.02.20	Demo for conducting inference on a LXMERT model with an FRCNN network to answer user-provided questions on user-provided images	LXMERT	Visual Question Answering	Hugging Face	CO Open in Colab
NER: End-to-End Example with Datasets and Transformers Scripts	12.02.20	Demo for training BERT on the XTREME multilingual benchmark for the NER task. Also uses Tensorboard for visualizing training	BERT	Named Entity Recognition (NER)	Manu Romero	CO Open in Colab
E2E-TensorflowTTS	12.02.20	Demo for conducting inference with realtime end to end text-to-speech using TensorflowTTS	Tacotron, FastSpeech and MELGAN	Text-to-Speech	TensorFlow	CO Open in Colab
Finetuning on Polyglot NER with Datasets and Transformers scripts for NER downstream task	12.02.20	Demo for training BERT on the Polyglot NER dataset. Also uses Tensorboard for visualizing training	BERT	Named Entity Recognition (NER)	Manu Romero	CO Open in Colab
End-to-End Masked Language Modeling with BERT	12.02.20	Demo for buildling a BERT model from scratch, train it with the masked language modeling task, and then finetune on a sentiment classification task	BERT	Language Modeling, Sentiment Analysis	Ankur Singh	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
BERT-Ranker for ClariQ	12.02.20	Demo for training BERT on the ClariQ dataset for ranking clarifying questions to queries	BERT	Ranking	Gustavo Penha	Open in Colab
Dialect Adapter	12.02.20	Demo for training Arabic Dialect Corpus on BERT Adapter for classifying Arabic by dialect	AraBERT	Classification	Nick Doiron	Open in Colab
NeMo Voice Swap Demo	12.02.20	Demo shows how to use NVIDIA NeMo swap a voice in the audio fragment with a computer generated one	QuartzNet, Tacotron2, Waveglow	Automatic Speech Recognition (ASR)	NVIDIA	Open in Colab
Pretrained Language Models for Downstream Tasks	12.02.20	Demo for initializing a transformer for NER using NeMo's framework	RoBERTa	Named Entity Recognition (NER)	NVIDIA	Open in Colab
Exploring NeMo NLP Tokenizers	12.02.20	Demo for initializing and using NeMo tokenizers	n/a	Tokenizer	NVIDIA	Open in Colab
Text Classification (Sentiment Analysis) with BERT	12.02.20	Demo for training and conducting inference with a transformer on the SST-2 dataset using NeMo's framework	BERT	Classification	NVIDIA	Open in Colab
Question Answering with SQuAD	12.02.20	Demo for training and conducting inference with a transformer on the SQuAD dataset using NeMo's framework	BERT	Question Answering	NVIDIA	Open in Colab
Token Classification (Named Entity Recognition)	12.02.20	Demo for training and conducting inference with a transformer on the GMB(Groningen Meaning Bank) using NeMo's framework	BERT	Named Entity Recognition (NER)	NVIDIA	Open in Colab
Joint Intent Classification and Slot Filling	12.02.20	Demo for training and conducting inference with a tranformer on a virtual assistant interaction dataset using NeMo's framework	BERT	Intent Classification	NVIDIA	Open in Colab
GLUE Benchmark	12.02.20	Demo for training a tranformer on the GLUE Benchmark using NeMo's framework	BERT	Classification, Regression	NVIDIA	Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Punctuation and Capitialization	12.02.20	Demo for training 2 token-level classifiers on top of the pretrained BERT model to predict punctuation and capitalization using NeMo's framework	BERT	Preprocessing	NVIDIA	Open in Colab
Named Entity Recognition - BioMegatron	12.02.20	Demo for training and conducting inference with BioMegatron on the NCBI Disease Dataset using NeMo's framework	BioMegatron	Named Entity Recognition (NER)	NVIDIA	Open in Colab
Relation Extraction - BioMegatron	12.02.20	Demo for training BioMegatron on the chemical-protein interactions (CHEMPROT) dataset using NeMo's framework	BioMegatron	Relation Extraction	NVIDIA	Open in Colab
Deep Learning for Developers with Ludwig	12.02.20	Demo for introducing the Ludgic framework and train the AG news dataset	n/a	Classification	Ludwig	Open in Colab
WebScraping Using BeautifulSoup	12.02.20	Demo for using Beautifulsoup for web scraping websites	Framework	Web Scraping	practical.ai	Open in Colab
Extracting Text From Images Tesseract	12.02.20	Demo for using tesseract for extracting text from images	Framework	OCR	practical.ai	Open in Colab
Tokenization Stemming Lemmatization Stopword Postagging	12.02.20	Demo for using SpaCy and NLTK for various preprocessing tasks	Framework	Preprocessing	practical.ai	Open in Colab
Data Augmentation using NLPaug	12.02.20	Demo for using nlpaug library for data augmentation such as character augmenter and word augmenter	Framework	Data Augmentation	practical.ai	Open in Colab
Snorkel	12.02.20	Demo for using the snorkel library for data labeling	Framework	Data Labeling	practical.ai	Open in Colab
One Hot Encoding of Text	12.02.20	Demo for using one hot-encoding of the NLTK library	Framework	Preprocessing	practical.ai	Open in Colab
Bag of Words	12.02.20	Demo for learning basic word representation processing via NLTK library	Framework	Preprocessing	practical.ai	Open in Colab
Bag of N-Grams	12.02.20	Demo for learning basic word representation processing via NLTK library	Framework	Preprocessing	practical.ai	Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
TF-IDF	12.02.20	Demo for using the TF-IDF representation of a document using SKlearn's TfidfVectorizer	Framework	Preprocessing	practical.ai	CO Open in Colab
Pre Trained Word Embeddings	12.02.20	Demo for using pre-trained word embedding models to represent text	Framework	Preprocessing	practical.ai	CO Open in Colab
Training Embeddings Using Gensim	12.02.20	Demo for training word embeddings on the wiki corpus by using Gensim and FastText	Framework	Preprocessing	practical.ai	CO Open in Colab
DocVectors using averaging Via SpaCy	12.02.20	Demo for creating a document vector via spaCy library	Framework	Preprocessing	practical.ai	CO Open in Colab
Training Dov2Vec Using Gensim	12.02.20	Demo for training a doc2vec model on a custom text via the Gensim library	Framework	Preprocessing	practical.ai	CO Open in Colab
Visualizing Embeddings Using TSNE	12.02.20	Demo for visualizing embeddings using TSNE from Sklearn	Framework	Visualization	practical.ai	CO Open in Colab
Visualizing Embeddings Using Tensorboard	12.02.20	Demo for using Tensorboard to visualize word embeddings	Framework	Visualization	practical.ai	CO Open in Colab
Data Agnostic RoBERTa-Based Natural Languageto SQL Query Generation	12.02.20	Demo for training a transformer for generating sql queries from natural language questions	RoBERTa	Text-to-SQL	Debaditya Pal	CO Open in Colab
GeDi-Guided GPT-2 XL	12.02.20	Demo for implementing GeDI on GPT-2 for text generation with specific prompts	GPT-2	Text Generation	Salesforce	CO Open in Colab
Using Tree-Hugger to Enhance CodeXGLUE	12.02.20	Demo for using Tree-Hugger library for creating your own code dataset	Framework	Preprocessing	Codist	CO Open in Colab
PERIN: Permutation-Invariant Semantic Parsing	12.02.20	Demo for conducting inference with PERIN with various graph frameworks	PERIN	Semantic Parsing	David Samuel & Milan Straka	CO Open in Colab
Leveraging Pre-Trained Language Model Checkpoints for Encoder-Decoder Models	12.02.20	Demo illustrating how to warm-start encoder-decoder models with Transformers	BERT2BERT	Summarization	Patrick Von Platen	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
KeyBERT-Hebrew	12.02.20	Demo for using the KeyBERT keyphrase extraction library on the Hebrew language	BERT	Keyphrase Extraction	bzliran	Open in Colab
Usage of TweetEval and Twitter-Specific RoBERTa Models	12.02.20	Demo for conducting inference on masked language modeling, tweet similarity and tweet classification tasks using Twitter-specific transformers	RoBERTa	Classification, STS, Langugage Modeling	Luis Espinosa	Open in Colab
Pysentimiento: A Transformer-Based Sentiment Analysis Library for Spanish	12.02.20	Demo showing how to conduct inference with pysentimiento, a sentiment analysis library for the Spanish language	Framework	Sentiment Analysis	Juan Manuel P�rez	Open in Colab
Final BERT Long Docs	12.02.20	Demo for training transformer with document sequence lengths larger than 250 using Keras and Tensorflow	BERT	Classification	Armand Olivares	Open in Colab
Training a Bert WithTabular Model for Clothing Review Recommendation Prediction	12.02.20	Demo for training multimodal-transformers for classification on tabular data	BERT	Classification	Georgian Partners	Open in Colab
BERT Demo Notebook	12.02.20	Demo for masked language completion and word error correction with Transformer in the Hebrew language using PyTorch	BERT	Masked Language Modeling	Eyal Gruss	Open in Colab
BERTSimilarity	12.02.20	Demo for conducting inference with BERT for sentence textual similarity task	BERT	Semantic Textual Similarity	Abhilash Majumder	Open in Colab
Fine-Tuning IndicBERT	10.10.20	Demo for training IndicBERT on the wiki-cloze dataset for the Gujarati language	IndicBERT	Masked LM	Tathagata Raha	Open in Colab
Silero Examples	10.10.20	Demo for conducting inference using speech-to-text models in TensorFlow, PyTorch and ONNX with the option of choosing to transcribe in English, German or Spanish language	Silero	Speech-to-Text	Alexander Veysov	Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Julia for Pythonistas	10.10.20	Demo for introducing the Julia programming language to beginners	n/a	Framework	Aurelien Geron	Open in Colab
Data Augmentation using NLPaug	10.10.20	Demo for introducing data augmentation techniques with the NLPaug library via a character augmenter and word augmenter	n/a	Data Augmentation	sai teja ponugoti	Open in Colab
GraphTSNE Demo	10.10.20	Demo for creating a visualization of the Cora graph dataset using GraphTSNE	n/a	Graph Visualization	Leow Yao Yang	Open in Colab
TableQA	09.26.20	Demo querying natural language on tabular data like csv or excel with BERT large trained on SQuAD dataset	BERT	Table Parsing	Abhijith Neil Abraham	Open in Colab
Text Generation with an RNN	09.26.20	Demo for preprocessing, training and generating text with an RNN on dataset of Shakespeare's writing	RNN	Text Generation	TensorFlow	Open in Colab
Customizing AdaNet With TensorFlow Hub Modules	09.26.20	Demo for conducting inference with AdaNet for sentiment analysis using TensorFlow Hub text embedding modules	AdaNet	Sentiment Analysis	TensorFlow	Open in Colab
Dataset Generation Tutorial	09.26.20	Demo for generating a training dataset for goal-oriented chatbot. Specifically, dialog from the dstc2 training dataset	n/a	Dataset Generation	DeepPavlov	Open in Colab
TensorflowTTS-TFLite	09.26.20	Demo for converting models from TensorflowTTS (Text-to-Speech) to TFlite	Framework	Model Conversion	TensorFlow	Open in Colab
DialogRPT	09.19.20	Demo for conducting inference with the DialogRPT model that predicts whether a reply will lead to more upvoting or higher engagement	GPT-2	Ranking	Xiang Gao	Open in Colab
Introduction: Hands-on Graph Neural Networks	09.19.20	Demo for introducing PyTorch Geometric library by learning the node embeddings of the Karate Club dataset	GCN	Framework	PyTorch Geometric	Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Node Classification	09.19.20	Demo for training a GCN in node classification on the Cora dataset, where nodes represent documents, using PyTorch Geometric	GCN	Node Classification	PyTorch Geometric	CO Open in Colab
Graph Classification	09.19.20	Demo for training a GCN in graph classification on the TUDatasets using PyTorch Geometric	GCN	Graph Classification	PyTorch Geometric	CO Open in Colab
Scaling GNNs	09.19.20	Demo for training a Cluster-GCN on the pubmed dataset using PyTorch Geometric	Cluster-GCN	Node Classification	PyTorch Geometric	CO Open in Colab
Questgen NLP Library Demo	09.11.20	Demo for using the QuestGen Library for question generation	T5	Text Generation	QuestGen	CO Open in Colab
Introduction to the TensorFlow Models NLP library	09.11.20	Demo for using the TF NLP Library for building common NLP tasks which includes: pretraining, span labelling and classification	BERT	Text Classification	TensorFlow	CO Open in Colab
Multi-task Training with Hugging Face Transformers and NLP	09.11.20	Demo for conducting multi-task training with the Hugging Face framework. Datasets used are STS-B, RTE and Commonsense QA.	RoBERTa	Multi-Task Training	Jason Phang	CO Open in Colab
T5 IMDB Yelp	09.11.20	Demo for training the T5 model on the IMDB and Yelp Polarity Review datasets using TensorFlow	T5	Sentiment Analysis	Yeshwanth Reddy	CO Open in Colab
Gradient Salience: Explaining the Output of a HuggingFace BertForQA model	09.11.20	Demo for reviewing the output of BERT Question Answering model in TensorFlow	BERT	Question Answering	Victor Dibia	CO Open in Colab
Data augmentation with transformer models for named entity recognition	09.06.20	Demo for using a pretrained masked language model to generate more samples for an NER dataset	DistilRoBERTa	Data Augmentation	Manu Romero	CO Open in Colab
Part 1: Introducing txtai	09.06.20	Demo for introducing and using txtai for running similarity searches and building an embeddings index	Framework	Search	David Mezzetti	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Part 2: Extractive QA with txtai	09.06.20	Demo for using txtai for extractive question answering by creating embedding and extractor instances	Framework	Question Answering	David Mezzetti	CO Open in Colab
Part 3: Build an Embeddings Index from a Data Source	09.06.20	Demo for reading data from SQLite, building a Embedding index backed by word embeddings and running queries against the generated Embeddings index	Framework	Search	David Mezzetti	CO Open in Colab
Part 4: Extractive QA with Elasticsearch	09.06.20	Demo for showing how to conduct extractive question-answering on top of an Elasticesearch system	Framework	Question Answering	David Mezzetti	CO Open in Colab
Zero Shot Pipeline	09.06.20	Demo for conducting inference with zero-shot classification with the Hugging Face pipeline framework	Framework	Zero-Shot Classification	Hugging Face	CO Open in Colab
Semantic Search Spanish	09.06.20	Demo for finding similar documents using faiss and transformers in the Spanish language	RuPERTa	Search	Manu Romero	CO Open in Colab
Mining Twitter	09.06.20	Demo for extracting elements via Twitter's API and also conducting sentiment analysis	API	Text Mining	Mikhail Klassen	CO Open in Colab
Mining LinkedIn	09.06.20	Demo for connecting to and extracting fields via LinkedIn's API and doing clustering analysis	API	Text Mining	Mikhail Klassen	CO Open in Colab
Transformer OCR	09.06.20	Demo for training and conducting inference for OCR using a Transformer	Transformer	OCR	Himanshu Chaudhary	CO Open in Colab
BERT for Text Extraction on DRCD dataset	09.06.20	Demo for training BERT on the DRCD Chinese dataset in TensorFlow	BERT	Question Answering	Vincent Wu	CO Open in Colab
Detect Fake News	09.06.20	Demo for preprocessing, training and conducting Inference with BERT for detecting fake news	BERT	Text Classification	Prafull Sharma	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Similarity News	09.06.20	Demo for preprocessing and training BERT sentence embeddings with TensorFlow for searching news articles based on textual similarity	BERT	Semantic Textual Similarity	Prafull Sharma	CO Open in Colab
Getting Started with PyTorch	07.14.20	Demo for introducing PyTorch and Numpy	n/a	Framework	Venelin Valkov	CO Open in Colab
Build Your First Neural Network with PyTorch	07.14.20	Demo for buildilng a neural network from scatch using PyTorch. You can build it, train it on a weather dataset and save it.	nn	Quantitative	Venelin Valkov	CO Open in Colab
Create a Dataset for Sentiment Analysis	07.14.20	Create a dataset for sentiment analysis by scraping user reviews for Android apps.	n/a	Web Scraping, Dataset Creation	Venelin Valkov	CO Open in Colab
Deploy BERT for Sentiment Analysis with Transformers by Hugging Face and FastAPI	07.14.20	Demo for deploying a sentiment analysis BERT with FastAPI	BERT	Deployment	Venelin Valkov	CO Open in Colab
Run Streamlit App From a Google Colab Notebook	07.14.20	Demo for running a streamlit frontend from Colab	n/a	Deployment	Manu Romero	CO Open in Colab
FARM Building Blocks	07.14.20	Demo for training BERT German on the GermEval 2018 dataset using the Deepset framework	BERT	Text Classification	Deepset	CO Open in Colab
How to Export Transformers Models to ONNX ?	07.14.20	Demo for converting a PyTorch or TensorFlow transformers model to ONNX and leverage onnxruntime to run inference tasks	n/a	Framework	Hugging Face	CO Open in Colab
Advanced Text Analysis with SpaCy and Scikit-Learn	07.14.20	Demo for using SpaCy and scikit-learn for several classical machine learning tasks	n/a	Various	Explosion	CO Open in Colab
SpaCy Introduction	07.14.20	Demo for using SpaCy for various text processing tasks: tokenizing, POS, NER, syntactic dependencies, chunking, embeddings	n/a	Framework	Explosion	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Pride & Prejudice Analysis	07.14.20	Demo for using SpaCy with the Pride & Prejudice corpus for extracting names of all the characters from the book, visualizing characters' occurences with regards to relative position in the book, automatically describing any character from the book, finding out which characters have been mentioned in a context of marriage, and build keywords extraction that could be used to display a word cloud	n/a	Framework	Explosion	CO Open in Colab
Topic Modeling with Gensim	07.14.20	Demo for preprocessing text with SpaCy and creating topic models w/ Gensim and visualizing them	n/a	Topic Modeling	Explosion	CO Open in Colab
Lightning Tour of SpaCy NLP Library	07.14.20	Demo for using SpaCy for various text processing tasks	n/a	Framework	Explosion	CO Open in Colab
EDA + Classifier + Streamlit on Palmer Penguin Dataset	07.14.20	Demo for doing exploratory data analysis and training a random forrest classifier, inference with streamlit support	Random Forrest	EDA, Classification	Manu Romero	CO Open in Colab
DataGene - Data Comparison Toolbox	07.14.20	Demo for using Datagene, a tool used to detect and compare dataset similarity between real and synthetic datasets as well as train, test, and validation datasets	n/a	Dataset Similarity	FirmAI	CO Open in Colab
PandaPy	07.14.20	Demo for introducing PandaPy, a tool developed to improve the usability of python for finance	n/a	Framework	FirmAI	CO Open in Colab
aitextgen Train a GPT-2 Text-Generating Model w/ GPU	07.14.20	Demo for training and conducting inference with GPT-2	GPT-2	Text Generation	Max Woolf	CO Open in Colab
Use Trainers from HuggingFace Transformers on GPUs for Multiple Choice	07.14.20	Demo for training (multiple-choice) SWAG dataset on either PyTorch or TensorFlow	BERT	Commonsense	Hugging Face	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Transformer TTS: A Text-to-Speech Transformer in TensorFlow 2	07.14.20	Demo for conducting inference with the Forward Transformer TTS and MelGAN models	Forward Transformer TTS, MelGAN	Text-to-Speech	Axel Springer	CO Open in Colab
Transformer TTS: A Text-to-Speech Transformer in TensorFlow 3	07.14.20	Demo for conducting inference with the Forward Transformer TTS and WaveRNN models	Forward Transformer TTS, WaveRNN	Text-to-Speech	Axel Springer	CO Open in Colab

Notebook	Added	Description	Model	Task	Creator	Source
Transformer TTS: A Text-to-Speech Transformer in TensorFlow 2	07.14.20	Demo for conducting inference with the Forward Transformer TTS and MelGAN models	Forward Transformer TTS, MelGAN	Text-to-Speech	Axel Springer	CO Open in Colab