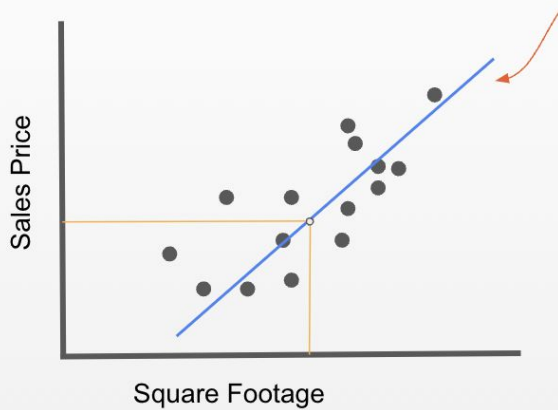


Large Scale Language Models and FinBERT

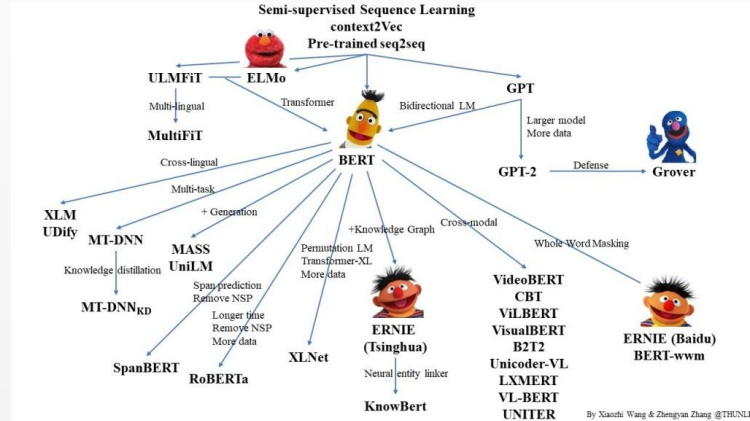
Rajiv Shah

raj@hf.com
@rajistics

Colab Notebook: https://bit.ly/raj_finbert



Supervised Machine Learning



Large Language Models

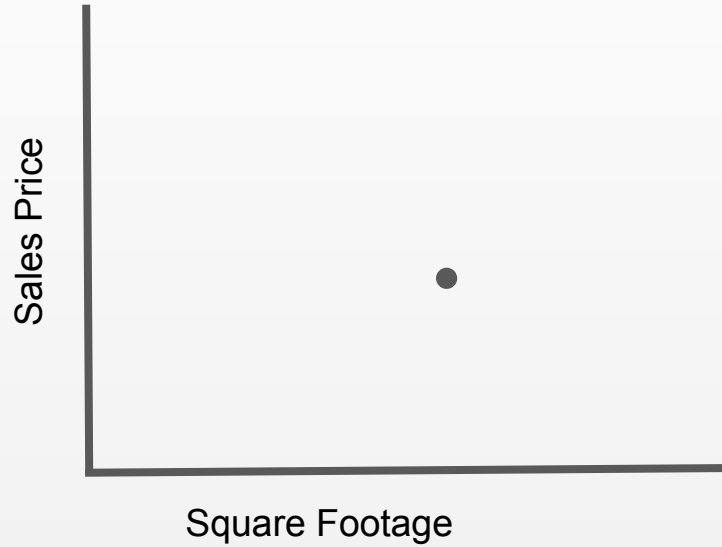


FinBERT

Colab Notebook: https://bit.ly/raj_finbert

Supervised Machine Learning

Machine Learning 1.0

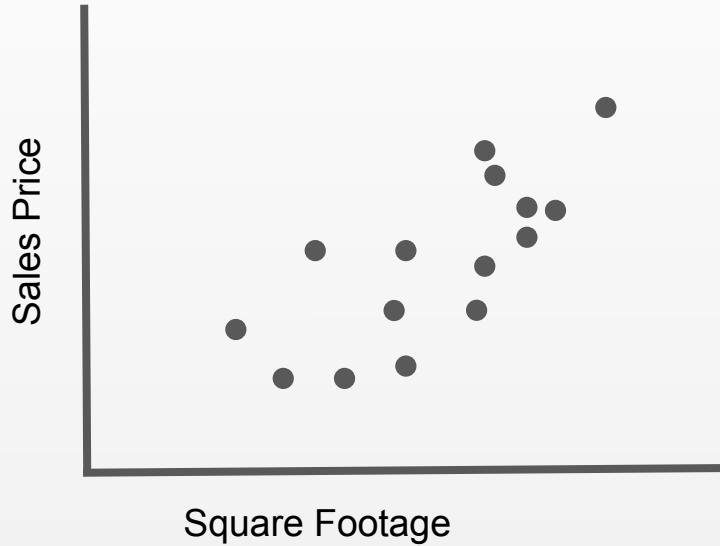


Machine Learning 1.0



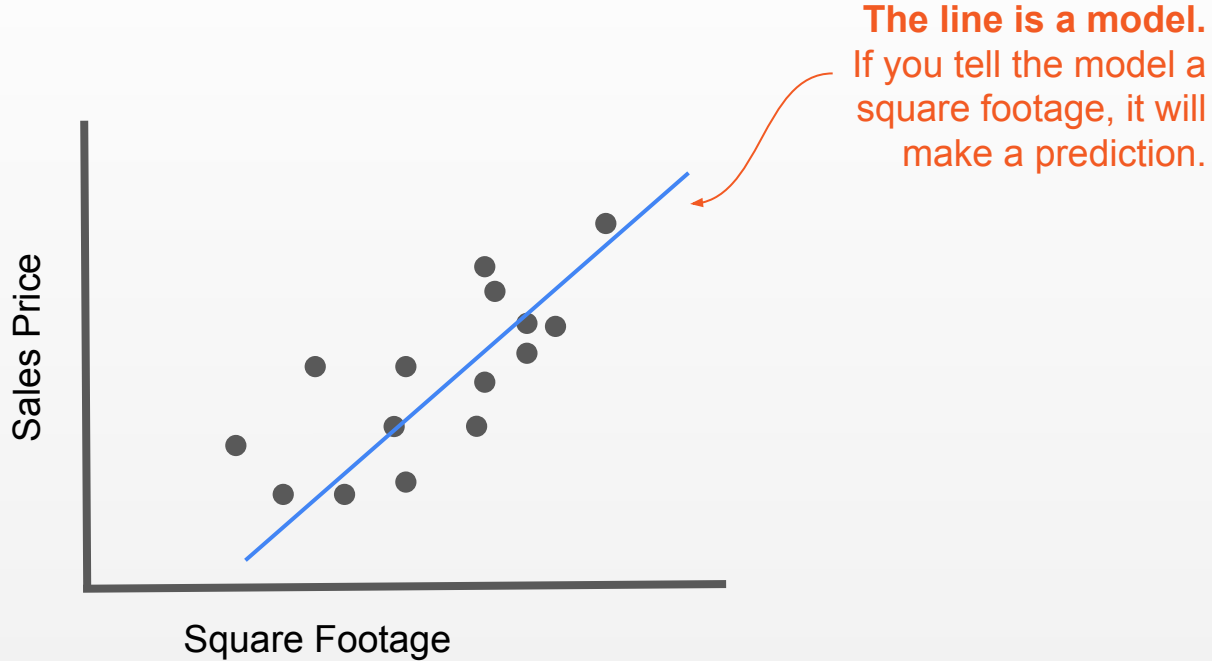
Every point is **an example** that the machine learns from

Machine Learning 1.0

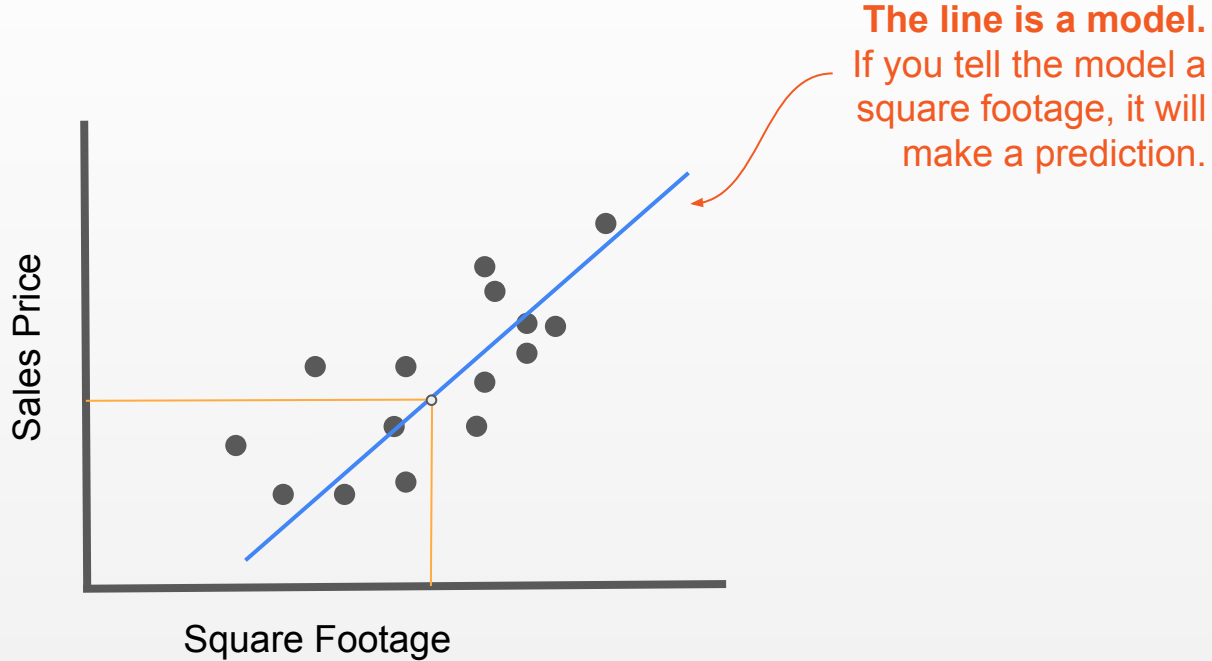


Every point is **an example** that the machine learns from

Machine Learning 1.0



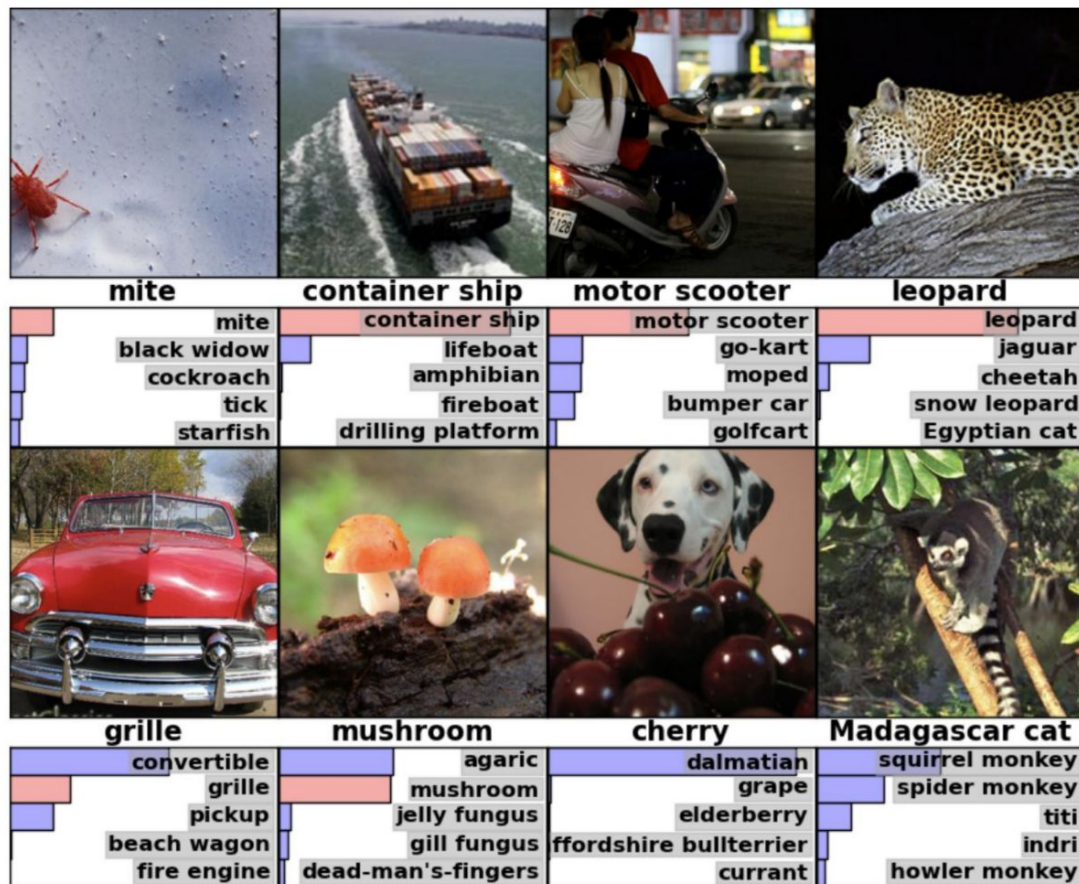
Machine Learning 1.0

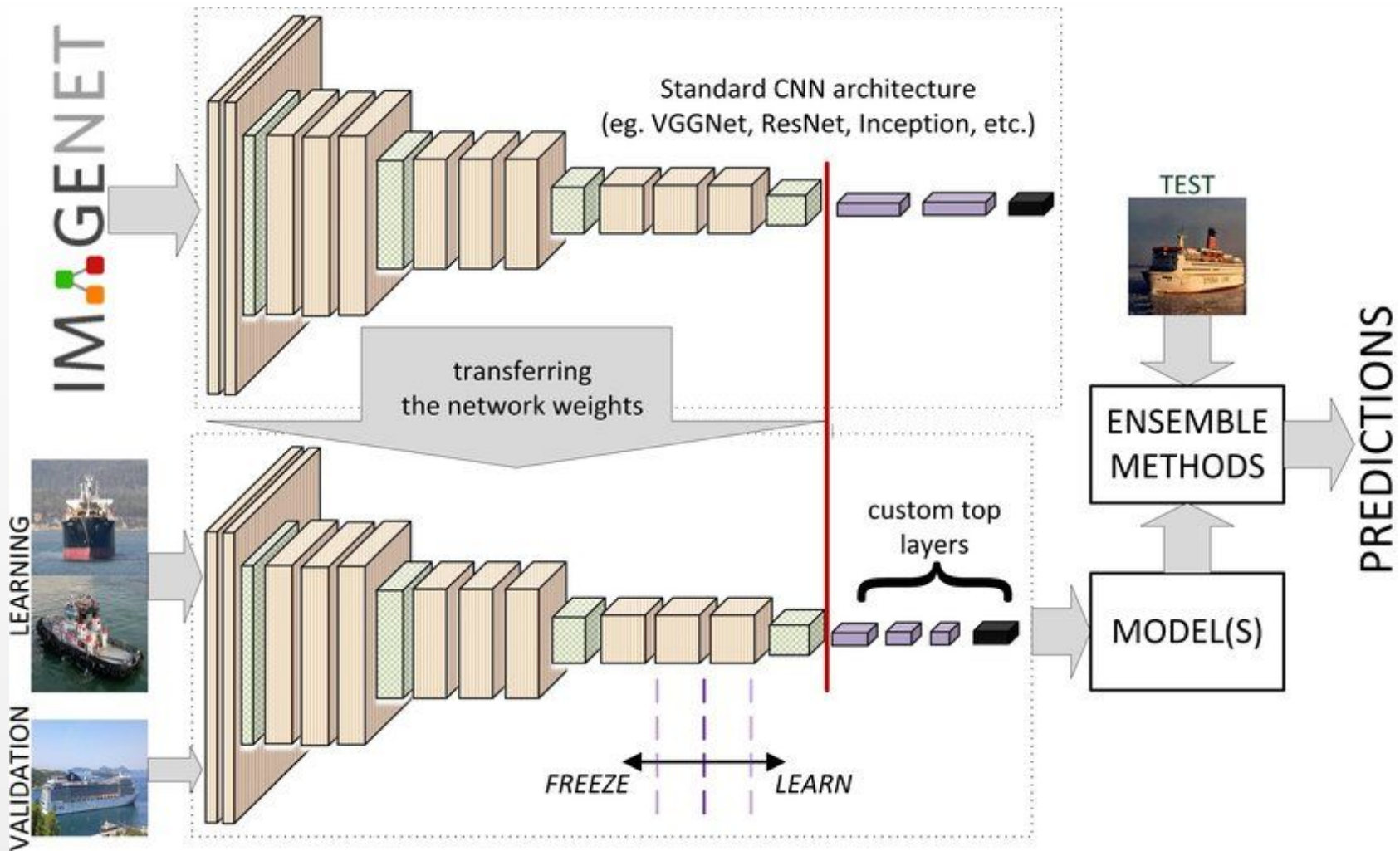


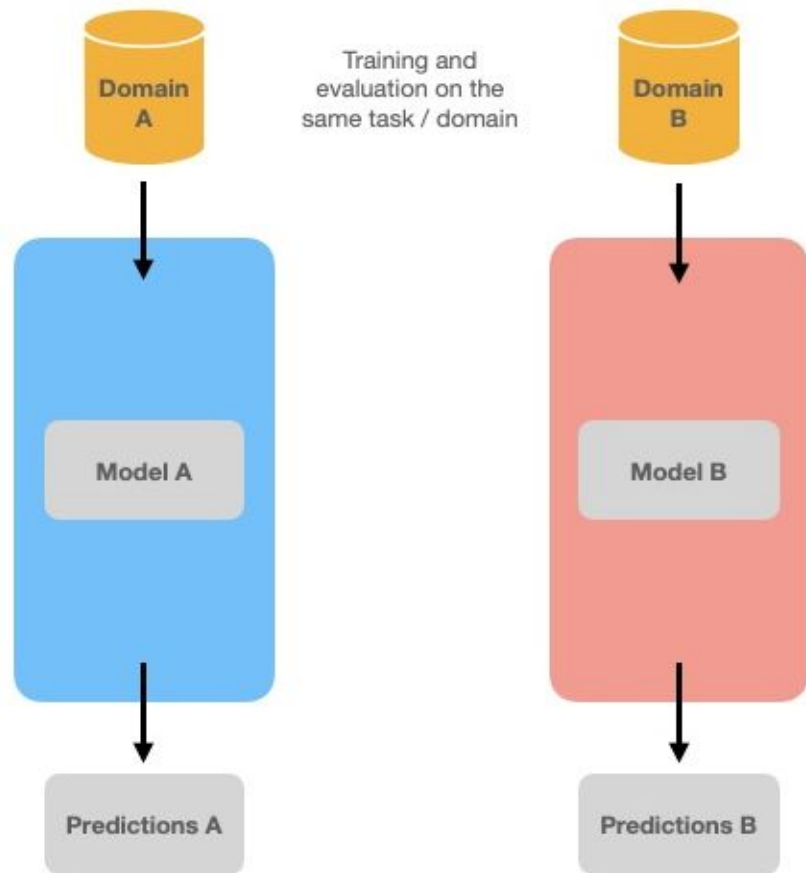
ImageNet Challenge



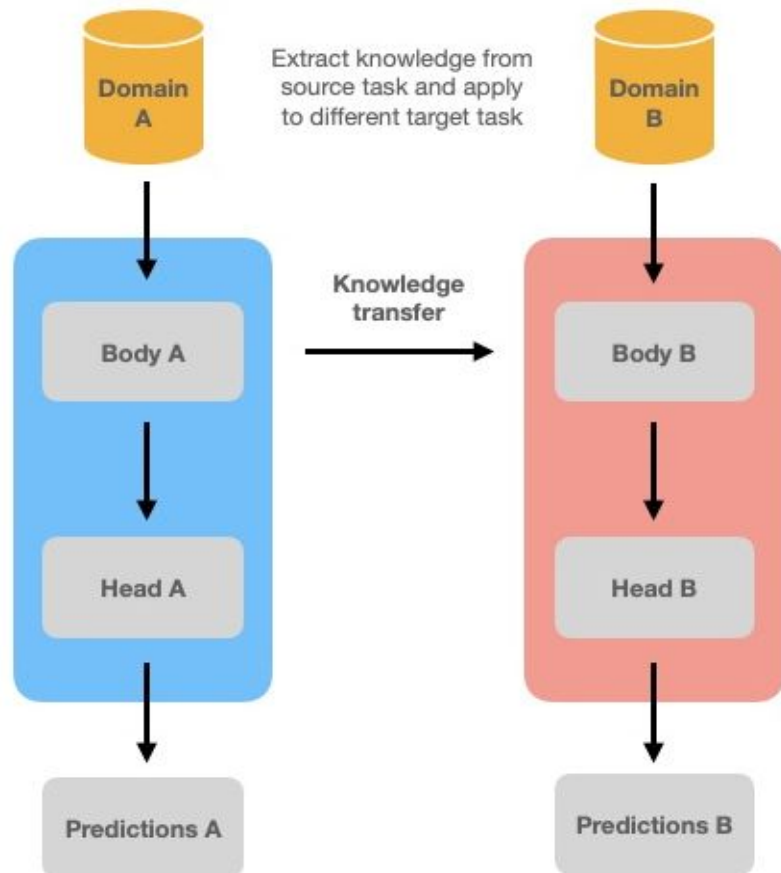
- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.













Supervised Learning






Transfer Learning


Let's explore some pretrained models at [hf.co](https://huggingface.co)

facebook/**wav2vec2-base-960h**   like 48

 Automatic Speech Recognition  PyTorch  TensorFlow  Transfor

[wav2vec2](#) [audio](#) [hf-asr-leaderboard](#) [Eval Results](#)



 **Model card**  Files  Community





 Edit r




Wav2Vec2-Base-960h


Facebook's Wav2Vec2

The base model pretrained and fine-tuned on 960 hours of Librispeech on 16kHz sampled speech audio. When using the model make sure that your speech input is also sampled at 16Khz.

facebook/**detr-resnet-50**   like 38

 Object Detection  PyTorch  Transformers  coco [arxiv:2005.12872](#)

 **Model card**  Files  Community

 Edit model

DETR (End-to-End Object Detection) model with ResNet-50 backbone

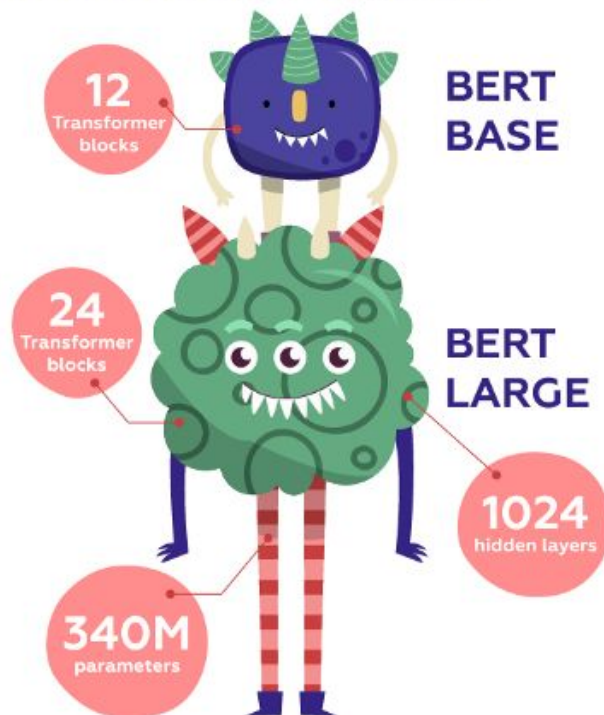
DEtection TRansformer (DETR) model trained end-to-end on COCO 2017 object detection (118k annotated images). It was introduced in the paper End-to-End Object Detection with Transformers by Carion et al. and first released in this repository.

Disclaimer: The team releasing DETR did not write a model card for this model so this model card has been written by the Hugging Face team.

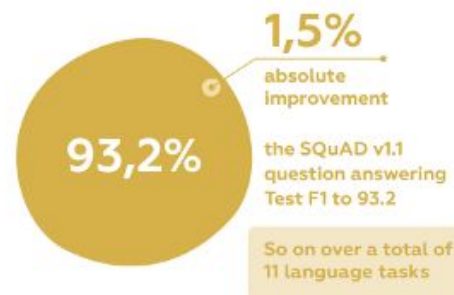
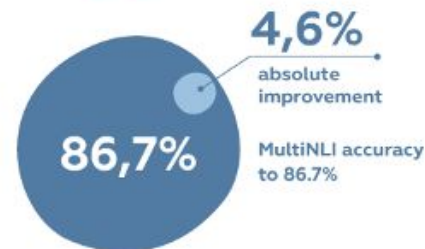
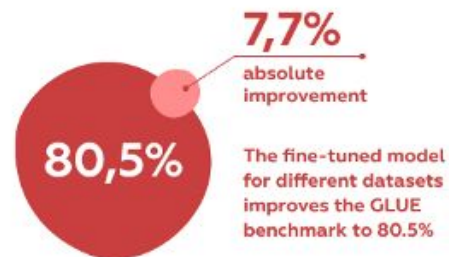
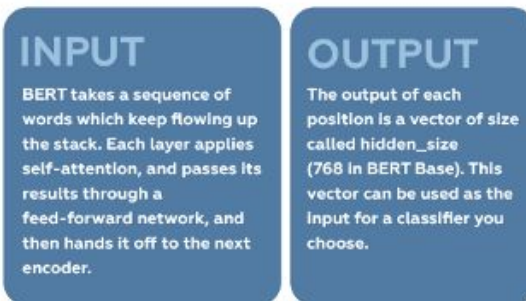
Large Language Models

BERT model at a glance

BERT comes in two sizes: BERT BASE, comparable to the OpenAI Transformer and BERT LARGE – the model which is responsible for all the striking results.



BERT is pre-trained on 40 epochs over:



Semi-supervised Sequence Learning

context2Vec
Pre-trained seq2seq



ELMo

ULMFiT

Multi-lingual

MultiFiT

Transformer

Bidirectional LM

GPT

Larger model
More data

GPT-2

Defense



Grover



BERT

Cross-lingual

Multi-task

+ Generation

XLM
UDify

MT-DNN

Knowledge distillation

MT-DNN_{KD}

MASS
UniLM

Span prediction
Remove NSP

Longer time
Remove NSP
More data

SpanBERT

RoBERTa

Permutation LM
Transformer-XL
More data

XLNet

+ Knowledge Graph



ERNIE
(Tsinghua)

Neural entity linker

KnowBert

Cross-modal

Whole Word Masking

VideoBERT
CBT




ViLBERT
VisualBERT
B2T2


Unicoder-VL
LXMERT
VL-BERT
UNITER




ERNIE (Baidu)
BERT-wwm


Rise of Transformers


   github.com/huggingface/transformers/releases?page=10



Nov 17, 2018

 thomwolf

 v0.1.2


 ce37b8e

Compare ▾

First release

This is the first release of `pytorch_pretrained_bert`.

► **Assets** 4





Let's use the transformer's library in the Colab Notebook

Transformer Pipelines

Natural Language Processing



Fill-Mask

3333 models



Question Answering

1852 models



Sentence Similarity

529 models



Summarization

452 models



Text Classification

7517 models



Text Generation

4987 models



Token Classification

2531 models



Translation

1590 models

https://huggingface.co/docs/transformers/main_classes/pipelines



Transformer Pipelines

Audio



Audio Classification

116 models



Audio-to-Audio

70 models



Automatic Speech Recognition

2551 models



Text-to-Speech

199 models

Computer Vision



Image Classification

529 models



Image Segmentation

60 models



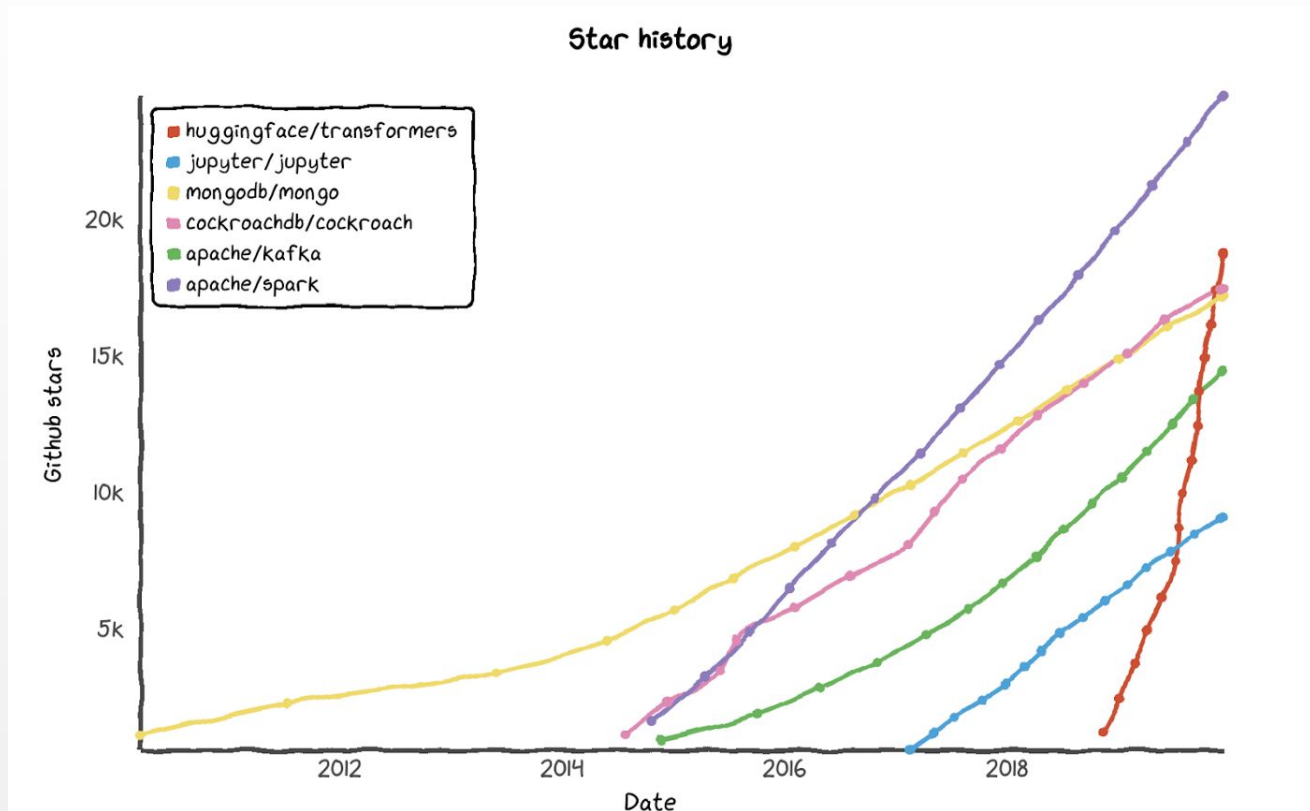
Object Detection

29 models

https://huggingface.co/docs/transformers/main_classes/pipelines

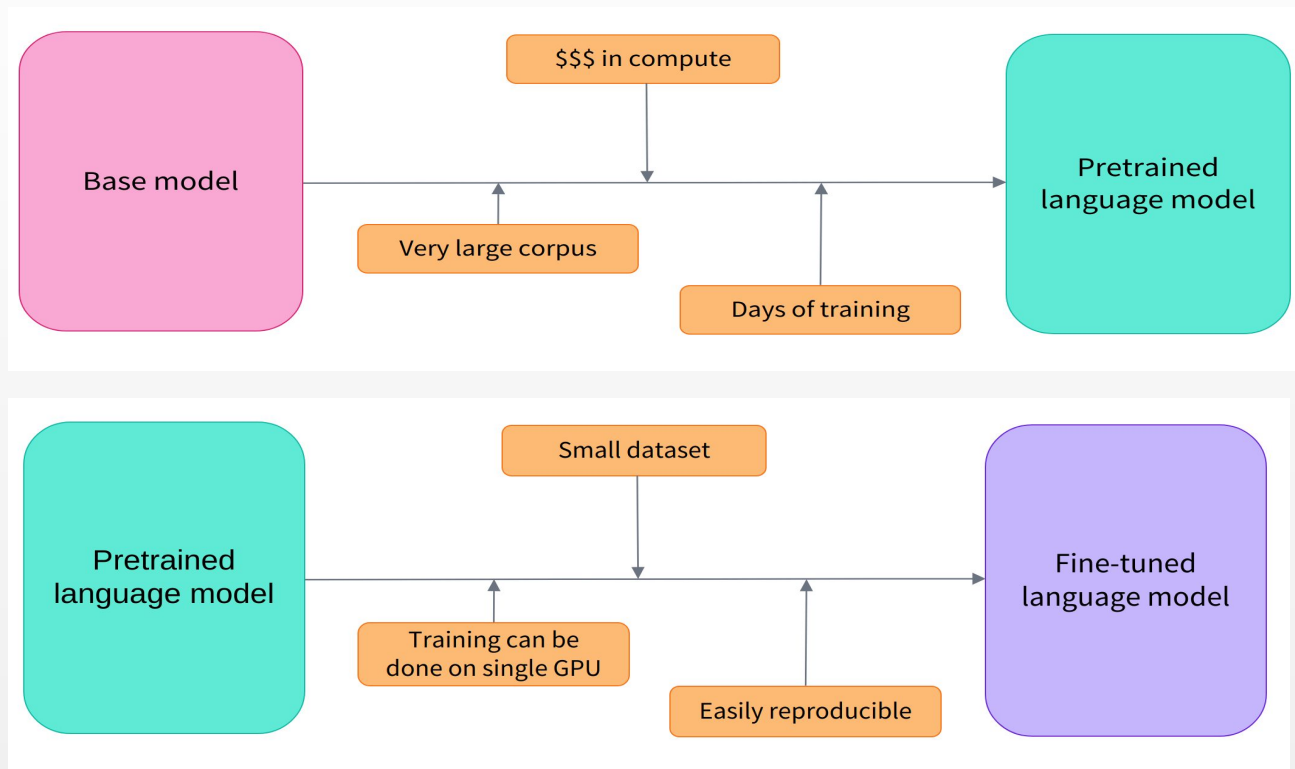


Rise of Transformers



FinBERT

Transfer Learning in NLP



FinBERT

Unlabelled financial text data

Reuters news corpus - 1.1 million words

Varity Corp, formerly Massey-Ferguson Ltd, said it expected to report on March 25 a loss for the fourth quarter and full-year ended January 31. A company spokesman said specific figures were unavailable.

Varity posted a net profit of 3.9 million U.S. dollars for the previous fiscal year ended January 31, 1986 and a 3.3 million dollar net profit for the previous fourth quarter.

Train



Result



Finance-specific
language model



FinBERT

Labelled financial text data

Financial PhraseBank - 5000 sentences classified by experts, with different agreement levels

Finnish Talentum reports its operating profit increased to EUR 20.5 mn in 2005 from EUR 9.3 mn in 2004 , and net sales totaled EUR 103.3 mn , up from EUR 96.4 mn .	+
Ford is struggling in the face of slowing truck and SUV sales and a surfeit of up-to-date , gotta-have cars .	-

Train



FinBERT

Negative

Score: -0.95

A second wave of layoffs amid weak demand and fractured supply chains is keeping new U.S. applications for unemployment benefits elevated

Positive

Score: 0.86

U.S. investors have ramped up investments in growth funds over the past few months on expectations of higher returns and safety.

Negative

Score: -0.45

The European Union said on Thursday it could impose taxes on digital giants such as Google, Amazon and Facebook even without a global agreement by the year-end.



FinBERT

"Such a massive asteroid hit will certainly create new business opportunities" said BERT. It's a new business opportunity, especially since asteroid impact is a bit of an unusual event, but there are still some risks to business.

Not everyone agrees with BERT though. ELMo, for example, is in disbelief of such an opportunistic view to a potential tragedy.

Many experts think that, if the asteroid hits Earth, it could affect our economy at a major cost and would increase our health care cost. And what about the impact on the planet? We're going to have to address that issue.

Analysis



Run

Export

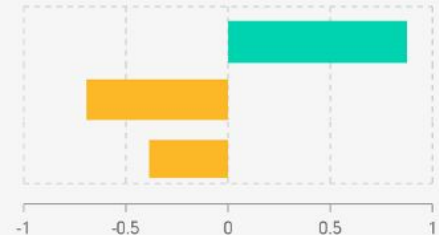
Characters: 588

Sentences: 7



CARDS

BARChart



FinBERT Performance

Model	All data			Data with 100% agreement		
	Loss	Accuracy	F1 Score	Loss	Accuracy	F1 Score
1. LSTM	0.81	0.71	0.64	0.57	0.81	0.74
2. LSTM with ELMo	0.72	0.75	0.7	0.50	0.84	0.77
3. ULMFit	0.41	0.83	0.79	0.20	0.93	0.91
4. LPS	-	0.71	0.71	-	0.79	0.80
5. HSC	-	0.71	0.76	-	0.83	0.86
6. FinSSLX	-	-	-	-	0.91	0.88
FinBERT	0.37	0.86	0.84	0.13	0.97	0.95

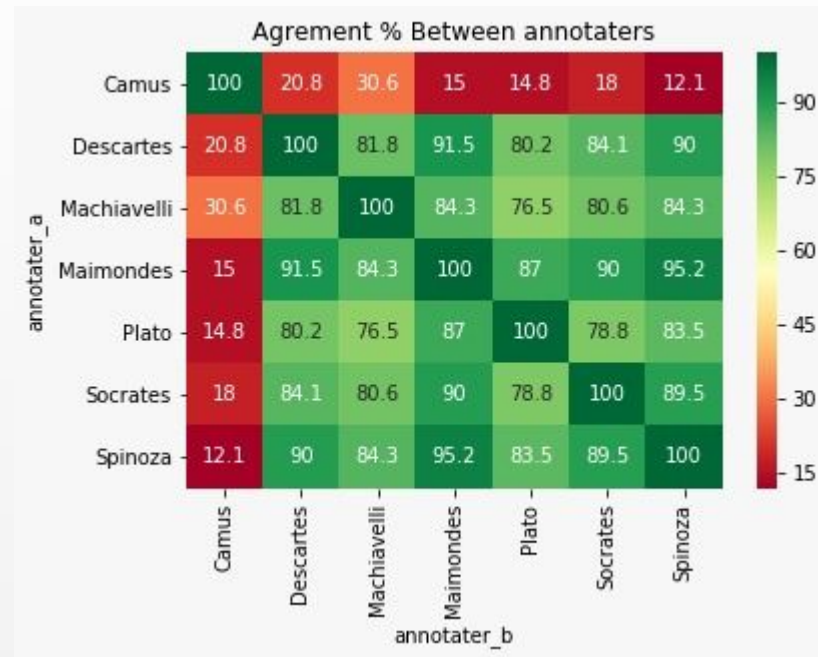


FinBERT - Annotator Agreement

Review annotator performance when assessing benchmarks

Identify inter-annotator disagreement

This will place a "ceiling" for any numbers we get. This will be useful when putting together a final error analysis.



FinBERT - Annotator Agreement

Positive-negative: 98.7%

Negative-neutral: 94.2%

Positive-neutral: 75.2%

It is because of the difficulty of distinguishing “commonly used company glitter and actual positive statements” and companies trying to spin an objectively neutral statement into a positive one.



FinBERT - Model Failures

A. Pre-tax loss totaled euro 0.3 million, compared to a loss of euro 2.2 million in the first quarter of 2005 .

True value: Positive

Predicted: Negative

This was a common failure model. The model sometimes fails to do the math in which figure is higher, and in the absence of words indicative of direction like “increased”, might make the prediction of neutral.



Large Scale Language Models and FinBERT

Rajiv Shah

raj@hf.com
@rajistics

Colab Notebook: https://bit.ly/raj_finbert