

1
oneAPI
<HACK>ATHON

BUILD SOLUTIONS TO
UNLOCK THE POTENTIAL OF
HETEROGENEOUS COMPUTING

LEAP powered by Intel® oneAPI AI Analytics Toolkit

Problem Statement : **Open Innovation in Education**

Team Name : C5ailabs

Team Members : Rohit Sroch, Sujith R Kumar, Mohan K Rachumallu, Shubham Jain

Problem Statement



MOOCs

(Massive Open Online Courses)

200K

Users in 2012

380M

Users in 2020

34.26%

CAGR 2022-27*

5% -10%

Completion rate

Key Challenges



Lengthy
videos



Instructor
Availability



Slow response
from forums



No real time
Q&A/Mentor

Approach

LEAP

(Learning Enhancement and Assistance Platform)



AI based
platform



Powered by
Intel OneAPI



Quality
Education



All time
Availability

Key Features of LEAP



Ask Question/Doubt



Conversational AI Examiner

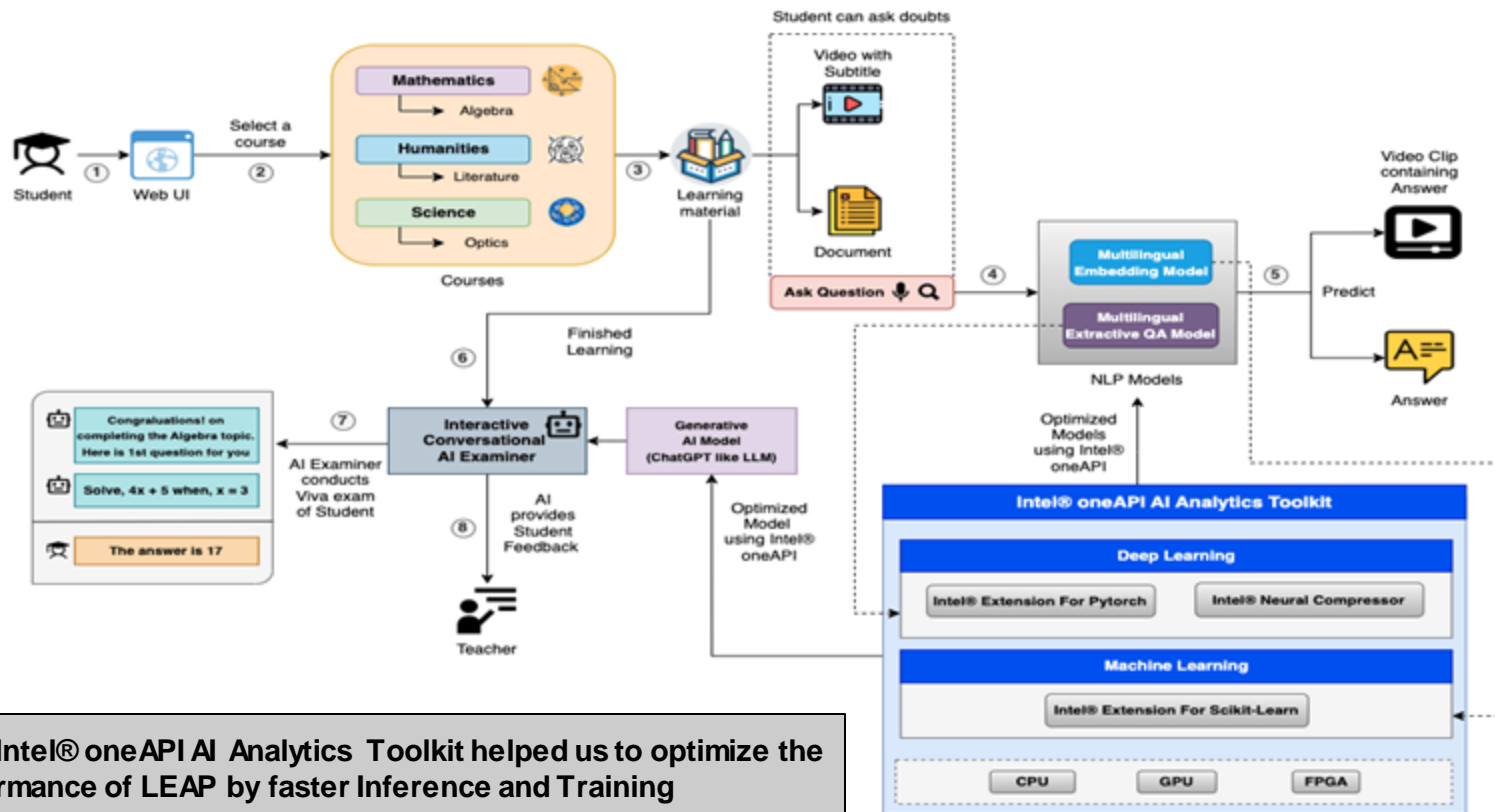


Feedback from AI Examiner



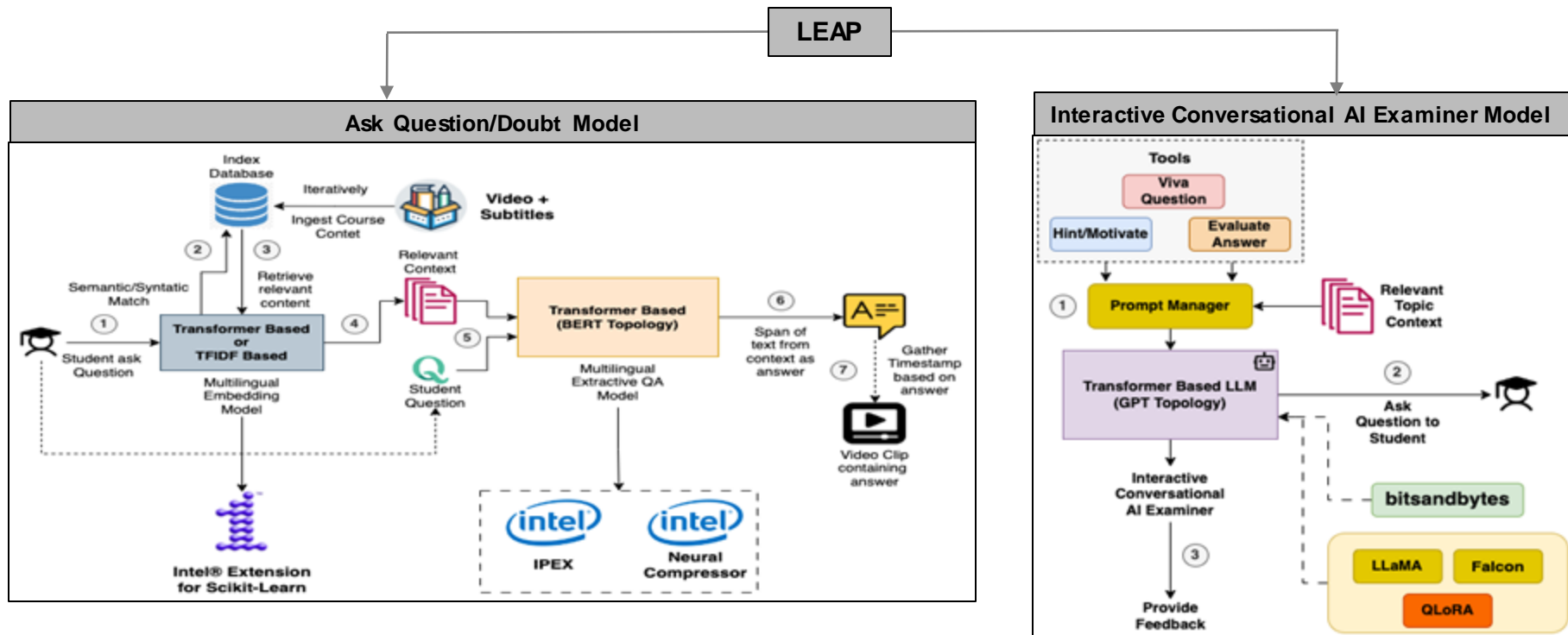
Multilingual Support

High Level Architecture

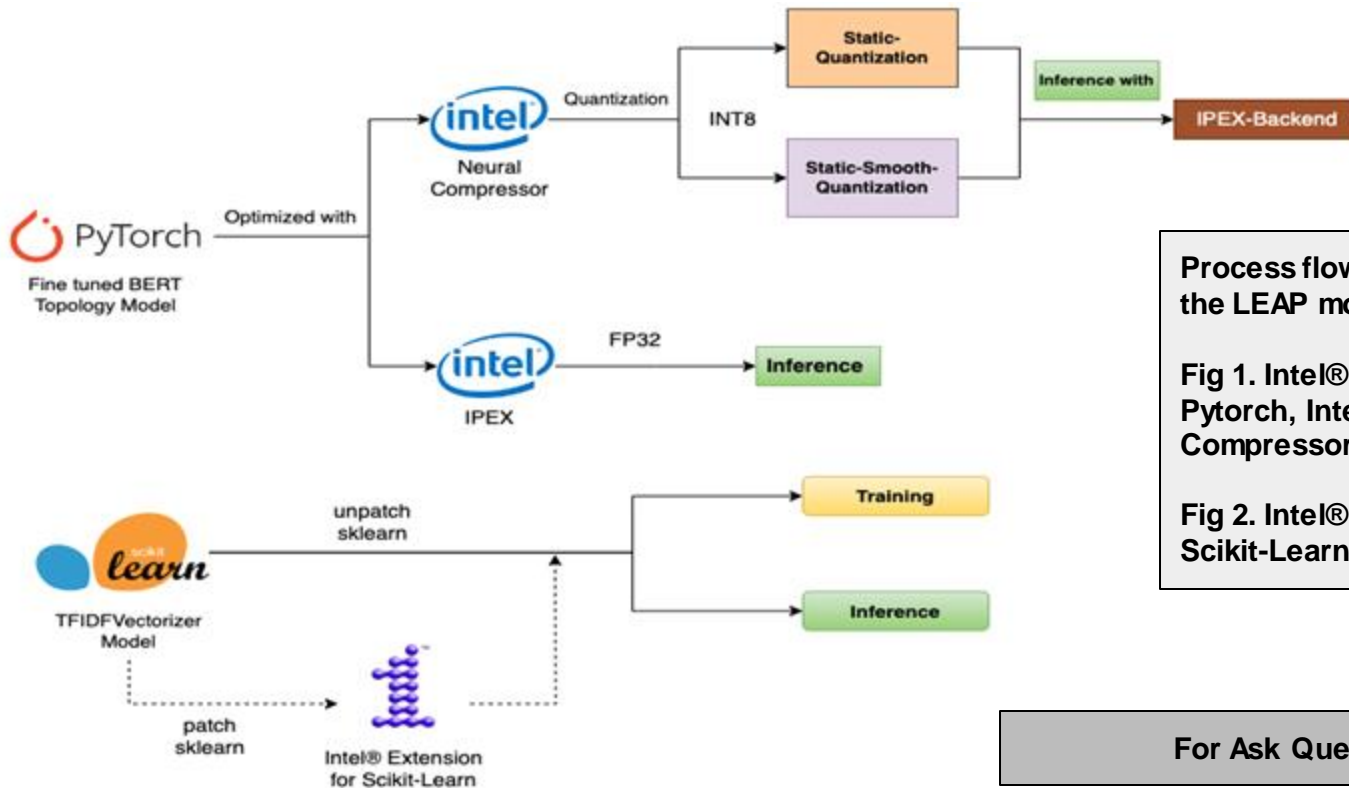


Here Intel® oneAPI AI Analytics Toolkit helped us to optimize the performance of LEAP by faster Inference and Training

LEAP: Detailed Model Architecture Diagram for Both Components



Result Summary (unique aspects of oneAPI/SYCL used)



Process flow to optimize the LEAP models by using

Fig 1. Intel® Extension for Pytorch, Intel® Neural Compressor and

Fig 2. Intel® Extension for Scikit-Learn

For Ask Question/Doubt

Demo Link and Screenshots

Link: <https://www.youtube.com/watch?v=CXkR5tklZm0>




Screen-1 (Dashboard Page)



Screen-2 (Week wise Content Page)

Demo Link and Screenshots




Course5 AI Labs
Learning The Future

Please note that this tool is only for demo purpose.

© Copyright 2023, Course5 AI Labs.

LEAP (Learning Enhancement and Assistance Platform)



Introduction to Deep Learning

Supervised Learning with Neural Networks

deeplearning.ai


0:00 / 0:00

Video Transcripts

00:00:00.000 -> 00:00:34.000

There's been a lot of hype about neural networks. And perhaps some of that hype is justified, given how well they're working. But it turns out that so far, almost all the economic value created by neural networks has been through one type of machine learning, called supervised learning. Let's see what that means, and let's go over some examples. In supervised learning, you have some input x , and you want to learn a function mapping to some output y . So for example, just now we saw the housing price prediction application where

Screen-3 (Ask Question/Doubt Page)



Course5 AI Labs
Shaping the Future

Please note that this tool is only for demo purpose.

© Copyright 2023, Course5 AI Labs.

what is ReLU

ReLU function which stands for rectified linear units.

Get More Info

Supervised Learning

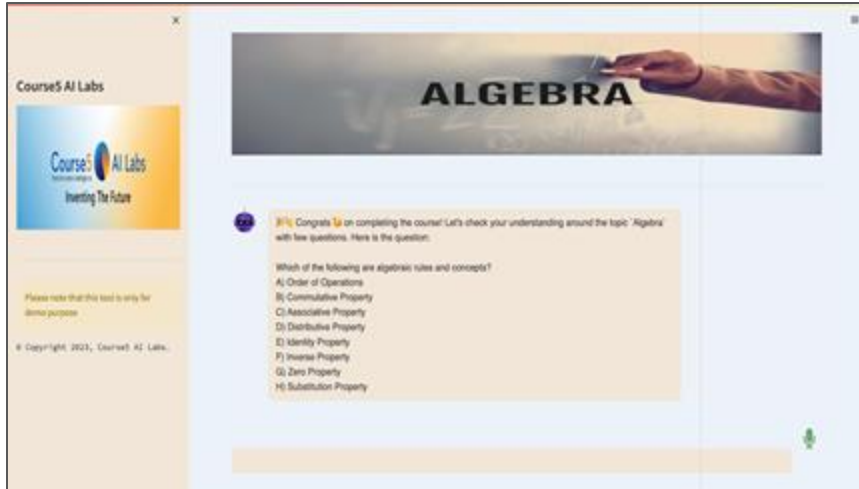
| Input(s) x^i | Output (y) \hat{y}^i | Application |
|----------------|------------------------|--------------------|
| Home features | Price | Real Estate |
| Ad, user info | Click or not CTR | Online Advertising |
| Image | Object (1, ..., 1000) | Photo tagging |

1:00 / 1:00

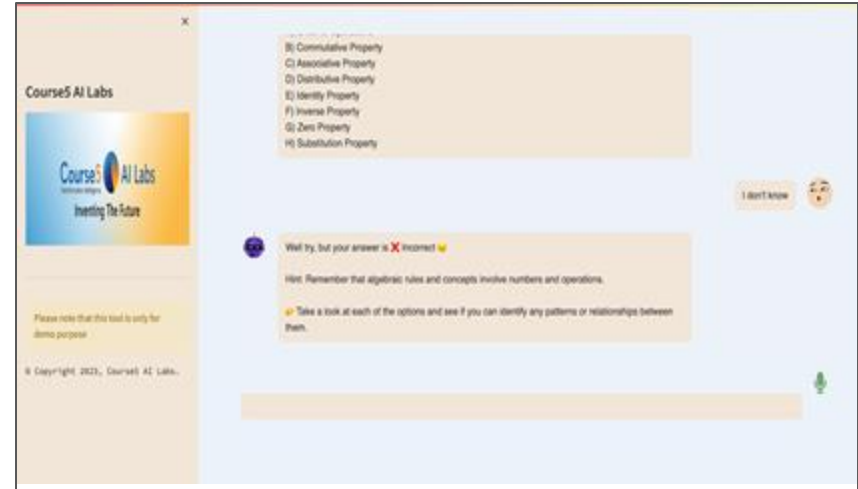
takes a row of data, and then outputs the estimated price. And by the way in the second column in literature, you can find this function a lot. This function which goes to zero whenever and then it's none of as a straight line. This function is called a ReLU function which stands for rectified linear units. So it's 1, 0. And mostly, just means taking a max of 0 which is why you get a function shape like this. You don't need to worry about ReLU units for now but it's just something you may encounter in this course.

Screen-4 (Ask Question/Doubt Page)

Demo Link and Screenshots



Screen-5 (Interactive Conversational AI Examiner Asks Question to student)



Screen-6 (Interactive Conversational AI Examiner provides hints and motivates a student in case of a wrong answer)

Extractive QA Model (BERT Topology) Latency/Speed-Up Comparison with IPEX and Intel® Neural Compressor

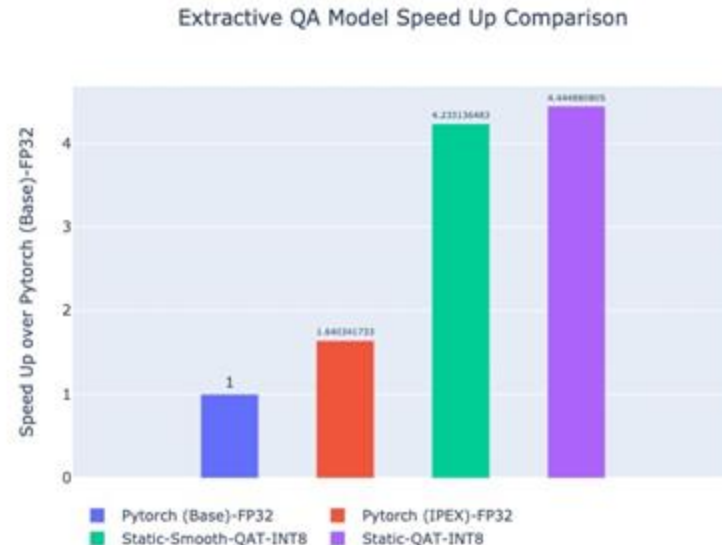
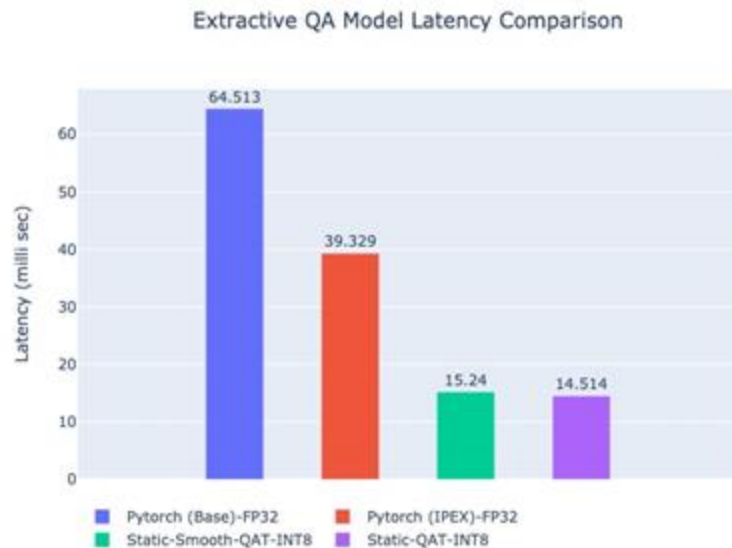
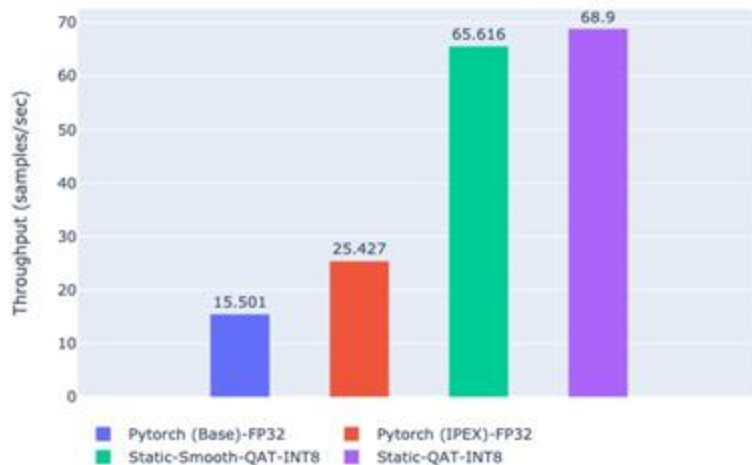


Fig: Latency/Speed-Up Benchmark result for **our Extractive Question Answering ALBERT Model (Multilingual)** on Intel® Dev Cloud machine (**Intel® Xeon® Platinum 8480+ (4th Gen: Sapphire Rapids) - 224v CPUs 503GB RAM**) with optimization using IPEX-FP32 and Static INT8-Quantization using Intel® Neural Compressor.

For Ask Question/Doubt Extractive QAModel

Extractive QA Model (BERT Topology) Throughput/F1 Score Comparison with IPEX and Intel® Neural Compressor

Extractive QA Model Throughput Comparison



Extractive QA Model F1 Score (SQuAD-v1) Comparison

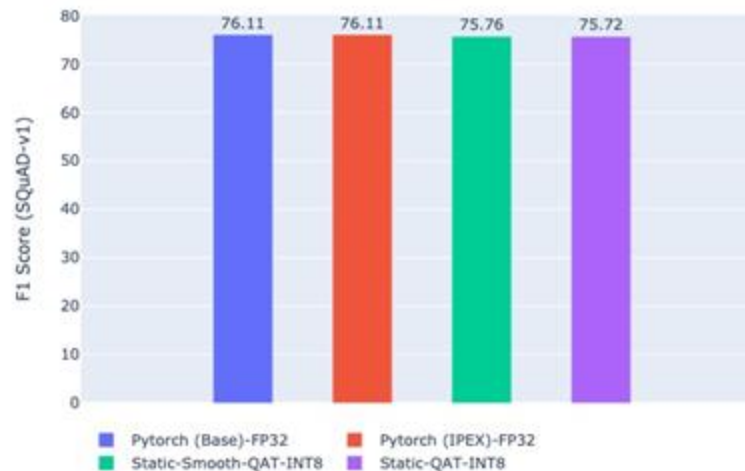


Fig: Throughput/F1 Score Benchmark result for **our Extractive Question Answering ALBERT Model (Multilingual)** on Intel® Dev Cloud machine (**Intel® Xeon® Platinum 8480+ (4th Gen: Sapphire Rapids) - 224v CPUs 503GB RAM**) with optimization using IPEX-FP32 and Static INT8-Quantization using Intel® Neural Compressor. Also, the model (<https://huggingface.co/ai4bharat/indic-bert>) was fine-tuned on SQuAD-v1 dataset.

For Ask Question/Doubt Extractive QAModel

Scikit-Learn (Base) vs Intel® Extension for Scikit-Learn

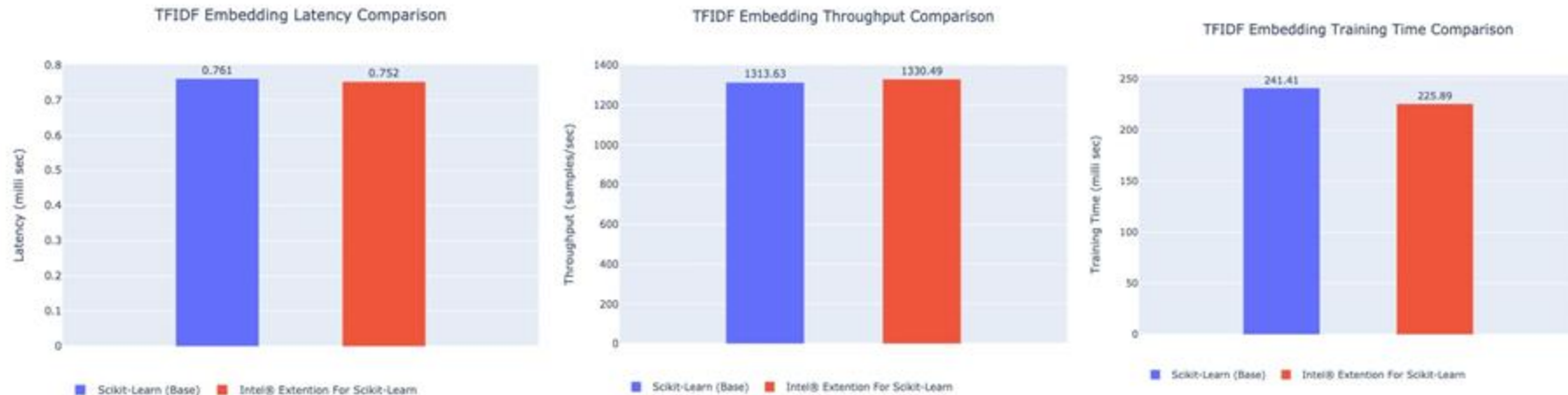


Fig: Benchmark results for **TFIDFVectorizer** Embedding model during training and inference on Intel® Dev Cloud machine (**Intel® Xeon® Platinum 8480+ (4th Gen: Sapphire Rapids) - 224v CPUs 503GB RAM**). Please Note that we don't see much of a difference may be because we used a tiny dataset.

GitHub Link (Codes should be public and available after hackathon also)

<https://github.com/rohitc5/intel-oneAPI>

| | | | | |
|--|-------------------------------------|--|--------------------|--------------|
| rohitc5 bump to latest based on feedback | | | dc59646 1 hour ago | 🕒 63 commits |
| api | bump to latest fixes | | | yesterday |
| assets | bug fixes | | | 4 hours ago |
| benchmark | upgrade to latest based on feedback | | | 8 hours ago |
| dataset | upgrade to latest | | | 3 weeks ago |
| nlp | upgrade to latest based on feedback | | | 8 hours ago |
| ppt | bump to latest based on feedback | | | 1 hour ago |
| webapp | bump to latest fixes | | | yesterday |
| .DS_Store | upgrade to latest | | | 3 weeks ago |
| LICENSE | upgrade to latest | | | 3 weeks ago |
| README.md | update README | | | 8 hours ago |
| docker-compose.yml | bump to latest | | | 3 weeks ago |

Step-by-Step Code Execution Instructions:

Quick Setup Option

- Make sure you have already installed docker (<https://docs.docker.com/get-docker/>) and docker-compose (<https://docs.docker.com/compose/>)
- Clone the Repository

```
$ git clone https://github.com/rohitc5/intel-oneAPI/  
$ cd intel-oneAPI
```

- Start the LEAP RESTFul Service to consume both components (Ask Question/Doubt and Interactive Conversational AI Examiner) as a REST API. Also Start the webapp demo build using streamlit.

```
# copy the dataset  
$ cp -r ../dataset webapp/  
  
# build using docker compose  
$ docker-compose build  
  
# start the services  
$ docker-compose up
```

- Go to <http://localhost:8502>

Model Checkpoint Release

<https://huggingface.co/rohitsroch>



The screenshot displays the Hugging Face profile of Rohit Sroch. On the left, the profile section includes a circular profile picture, the name 'Rohit Sroch', the username 'rohitsroch', and buttons for 'Edit profile' and 'Settings'. Below this, social media links for Twitter and GitHub are shown, followed by 'Research interests' (Speech Recognition, NLP) and 'Organizations' (one organization listed with a logo).

The main section is titled 'Models' with a sub-header '7' and a sort button 'Sort: Recently Updated'. It lists five models in a two-column grid:

- Model 1:** `rohitsroch/indic-mALBERT-static-smooth-INT8-s...`
Task: Question Answering - Updated about 5 hours ago
- Model 2:** `rohitsroch/indic-mALBERT-static-INT8-squad-v2`
Task: Question Answering - Updated about 5 hours ago
- Model 3:** `rohitsroch/indic-mALBERT-squad-v2`
Task: Question Answering - Updated about 8 hours ago · ⬆ 25
- Model 4:** `rohitsroch/hybrid_hbh_t5-small_ami_sum`
Task: Text2Text Generation - Updated Jun 13, 2022 · ⬆ 14
- Model 5:** `rohitsroch/hybrid_hbh_bart-base_icsi_sum`
Task: Text2Text Generation - Updated Jun 13, 2022 · ⬆ 13
- Model 6:** `rohitsroch/hybrid_utt-clusterrank_bart-base_s...`
Task: Text2Text Generation - Updated Mar 23 · ⬆ 10
- Model 7:** `rohitsroch/hybrid_utt-clusterrank_bart-base_d...`
Task: Text2Text Generation - Updated Jun 13, 2022 · ⬆ 39

Below the models list, the 'Datasets' section is shown with the text 'None yet'.

1
oneAPI
<HACK>ATHON

THANK YOU