LEAP powered by Intel® oneAPI AI Analytics Toolkit

Problem Statement : **Open Innovation in Education**

Team Name : C5ailabs

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

Problem Statement

- MOOCs (Massive Open Online Courses) have surged in popularity in recent years, particularly during the COVID-19 pandemic. These online courses are typically free or low-cost, making education more accessible worldwide.

- Online learning is crucial for students even post-pandemic due to its flexibility, accessibility, and quality. But still, the learning experience for students is not optimal, as in case of doubts they need to repeatedly go through videos and documents or ask in the forum which may not be effective because of the following **challenges**:

  - Resolving doubts can be a time-consuming process.
  - It can be challenging to sift through pile of lengthy videos or documents to find relevant information.
  - Teachers or instructors may not be available around the clock to offer guidance

- To mitigate the above challenges, we propose **LEAP (Learning Enhancement and Assistance Platform)**, which is an AI-powered platform designed to enhance student learning outcomes and provide equitable access to quality education. The platform comprises two main features that aim to improve the overall learning experience of the student:

❑ **Ask Question/Doubt:** This allows the students to ask real-time questions around provided reading material, which includes videos and documents, and get back answers along with the exact timestamp in the video clip containing the answer (so that students don't have to always scroll through). Also, It supports asking multilingual question, ensuring that language barriers do not hinder a student's learning process.

❑ **Interactive Conversational AI Examiner**: This allows the students to evaluate their knowledge about the learned topic through an AI examiner conducting viva after each learning session. The AI examiner starts by asking question and always tries to motivate and provide hints to the student to arrive at correct answer, enhancing student engagement and motivation.

List of features offered by the solution

Our proposed **LEAP (Learning Enhancement and Assistance Platform)** is an AI-powered platform designed to enhance student learning outcomes and provide equitable access to quality education. The platform comprises two main features that aim to improve the overall learning experience of the student:

- **Ask Question/Doubt:** This allows the students to ask real-time questions around provided reading material, which includes videos and documents, and get back answers along with the exact timestamp in the video clip containing the answer (so that students don't have to always scroll through). Also, It supports asking multilingual question, ensuring that language barriers do not hinder a student's learning process.

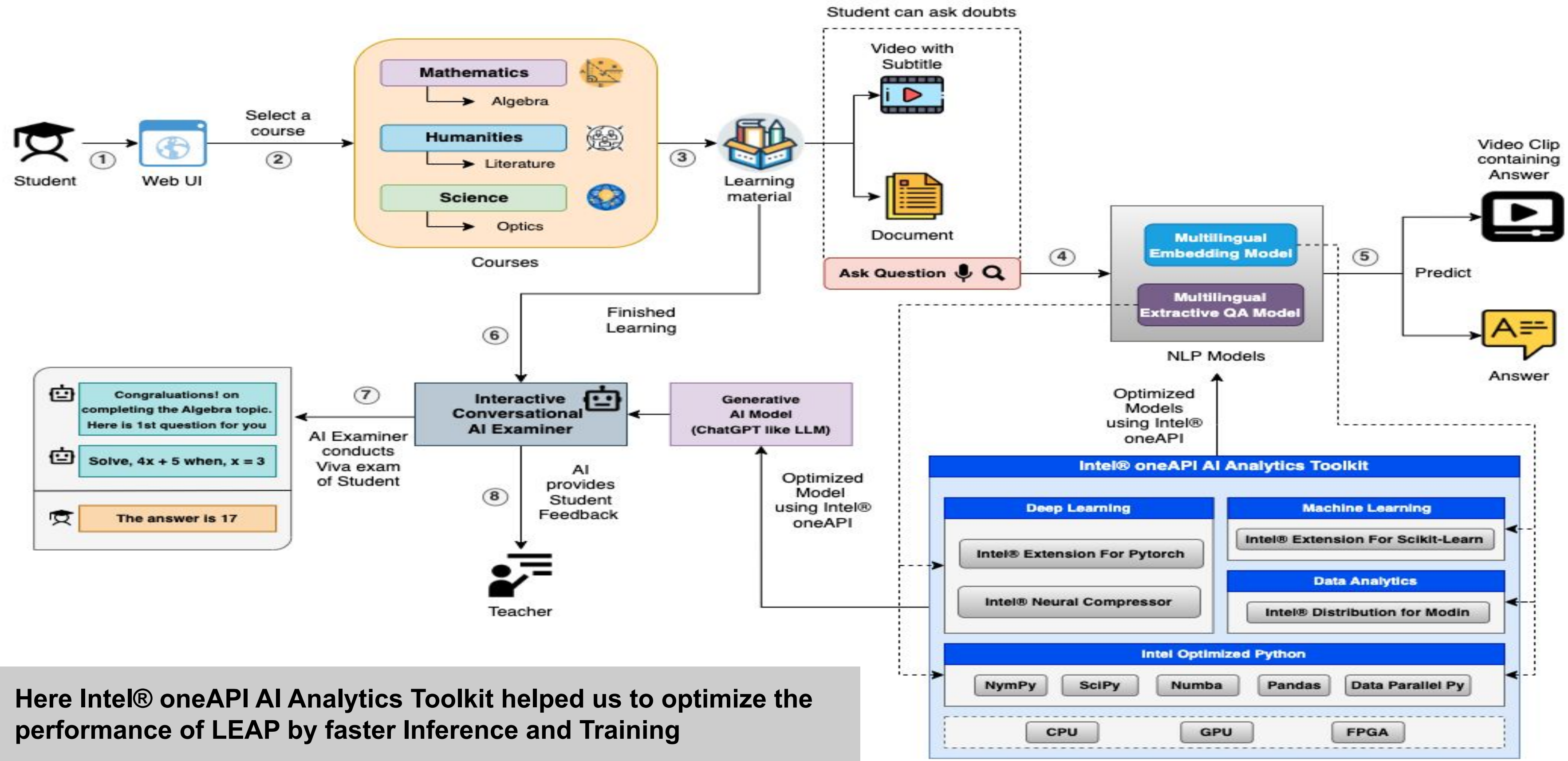
- **Interactive Conversational AI Examiner**: This allows the students to evaluate their knowledge about the learned topic through an AI examiner conducting viva after each learning session. The AI examiner starts by asking question and always tries to motivate and provide hints to the student to arrive at correct answer, enhancing student engagement and motivation.

Apart from above two main feature, following are a few more features

- **Feedback from AI Examiner:** Based on the performance of the student in the viva exam, the AI examiner can provide feedback of the student to the teacher which helps the teacher to pay attention to the weak points of student and prioritize accordingly.

- **Highly Reliable:** LEAP provides its answer by extracting the span of text from context as answer and can be trained specifically on the edu course content, ensuring that its answers are accurate and reliable.

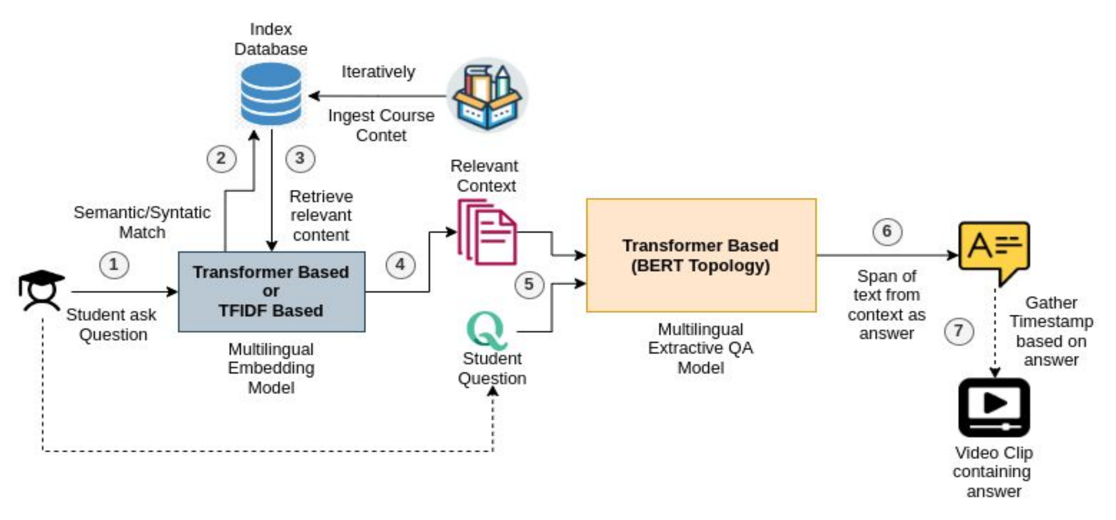
High Level Architecture



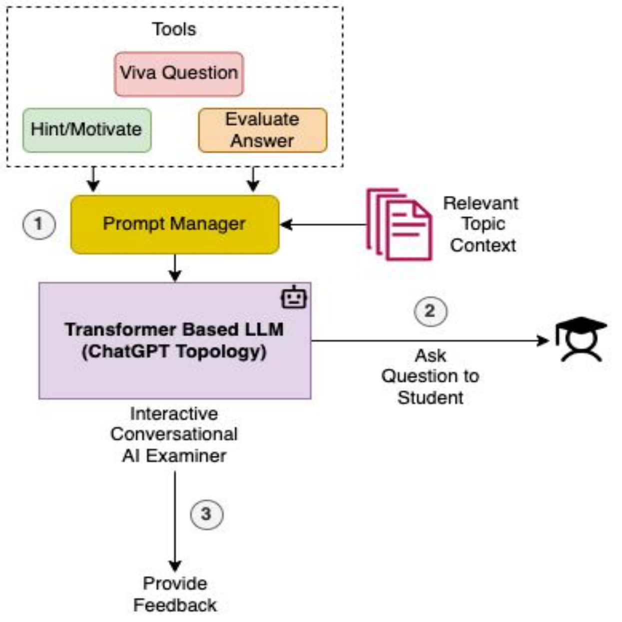
Here Intel® oneAPI AI Analytics Toolkit helped us to optimize the performance of LEAP by faster Inference and Training

Detailed Model Architecture Diagram for Both components of our LEAP

Ask Question/Doubt Model

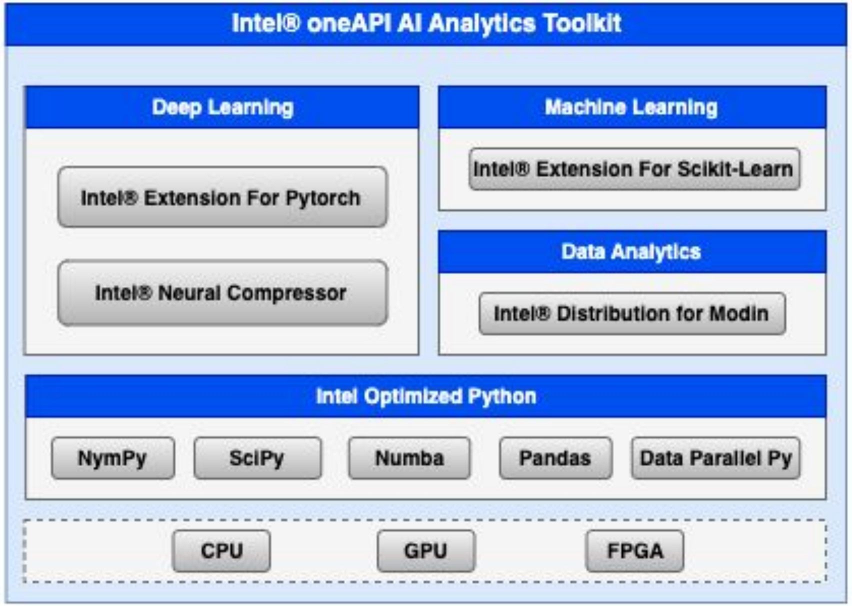


Interactive Conversational AI Examiner Model



Core components of oneAPI/SYCL used in the project

- Intel® Extension for Pytorch: Used for our Multilingual Extractive QA model Training/Inference optimization.
- Intel® Neural Compressor: Used for Multilingual Extractive QA model and Generative AI model Inference optimization.
- Intel® Extension for Scikit-Learn: Used for Multilingual Embedding model Training/Inference optimization.
- Intel® distribution for Modin: Used for basic initial data analysis.
- Intel® optimized Python: Used for data pre-processing, reading etc.



Mainly we have showed benchmark results with Intel® Extension for Pytorch, Intel® Neural Compressor and Intel® Extension for Scikit-Learn

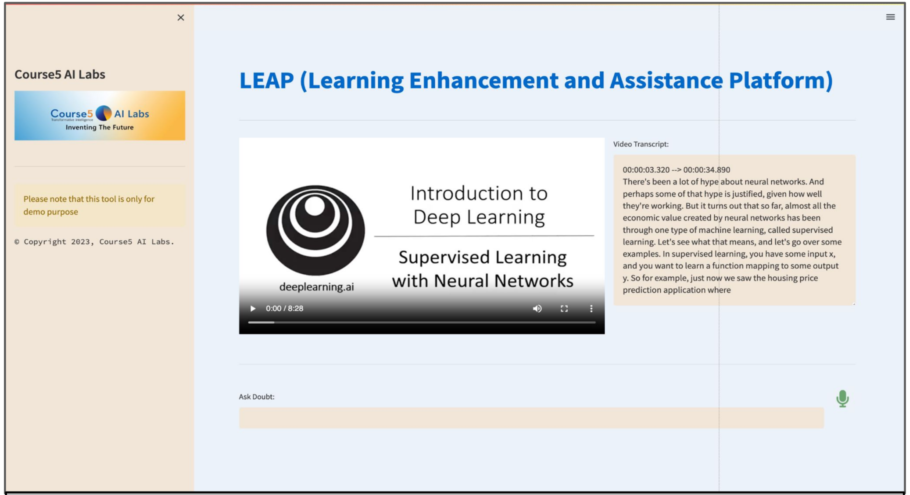
Demo Link and Screenshots

Link: https://www.youtube.com/watch?v=QoVWsOSlwvI

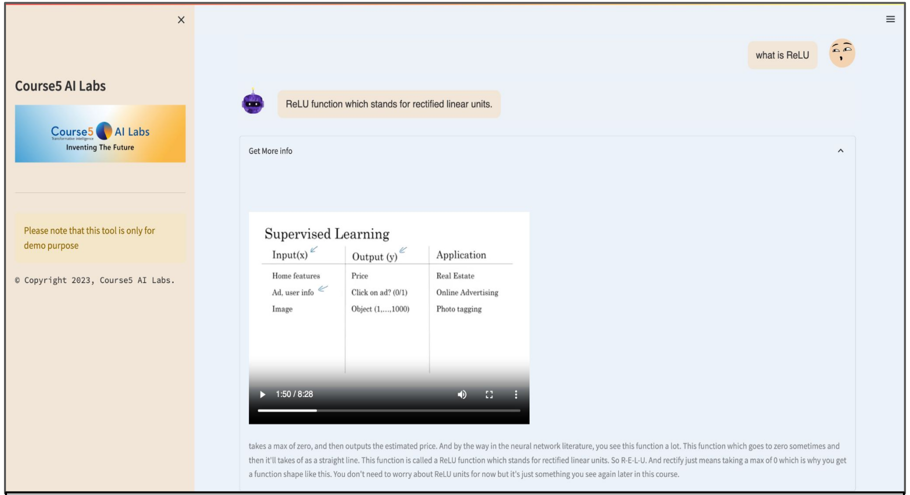
Screen-1 (Dashboard Page)

Screen-2 (Week wise Content Page)

Demo Screenshots

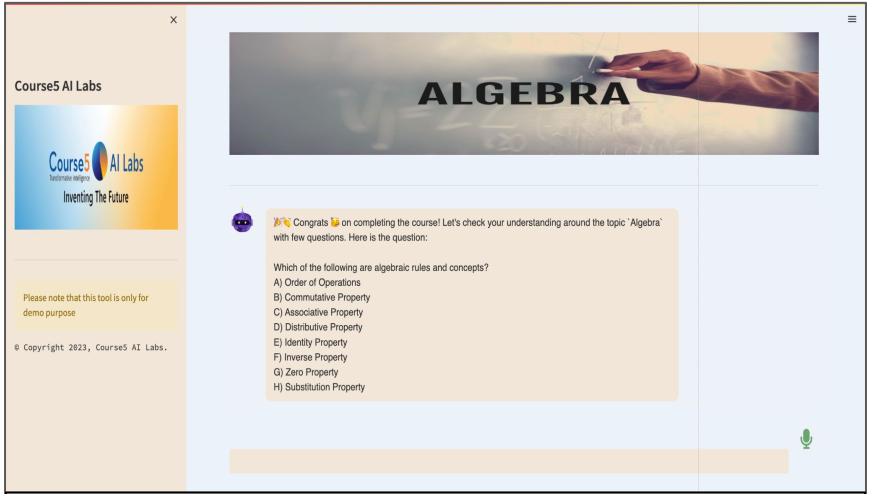


Screen-3 (Ask Question/Doubt Page)

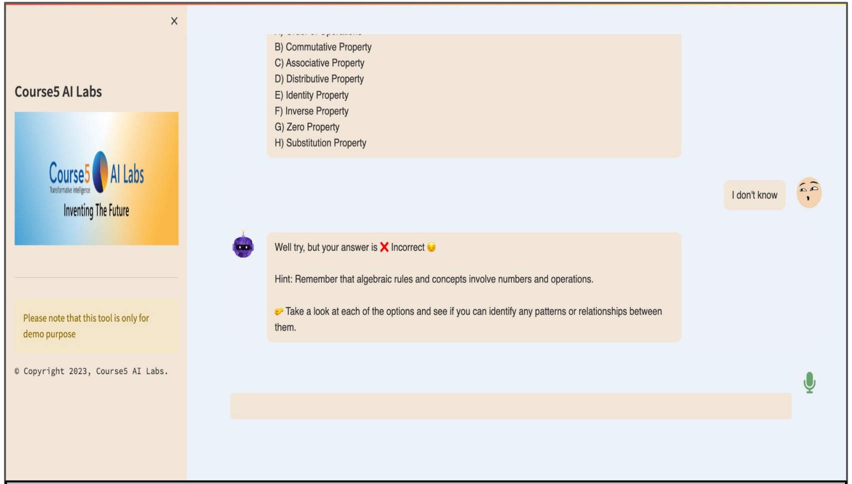


Screen-4 (Ask Question/Doubt Page)

Demo Screenshots



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

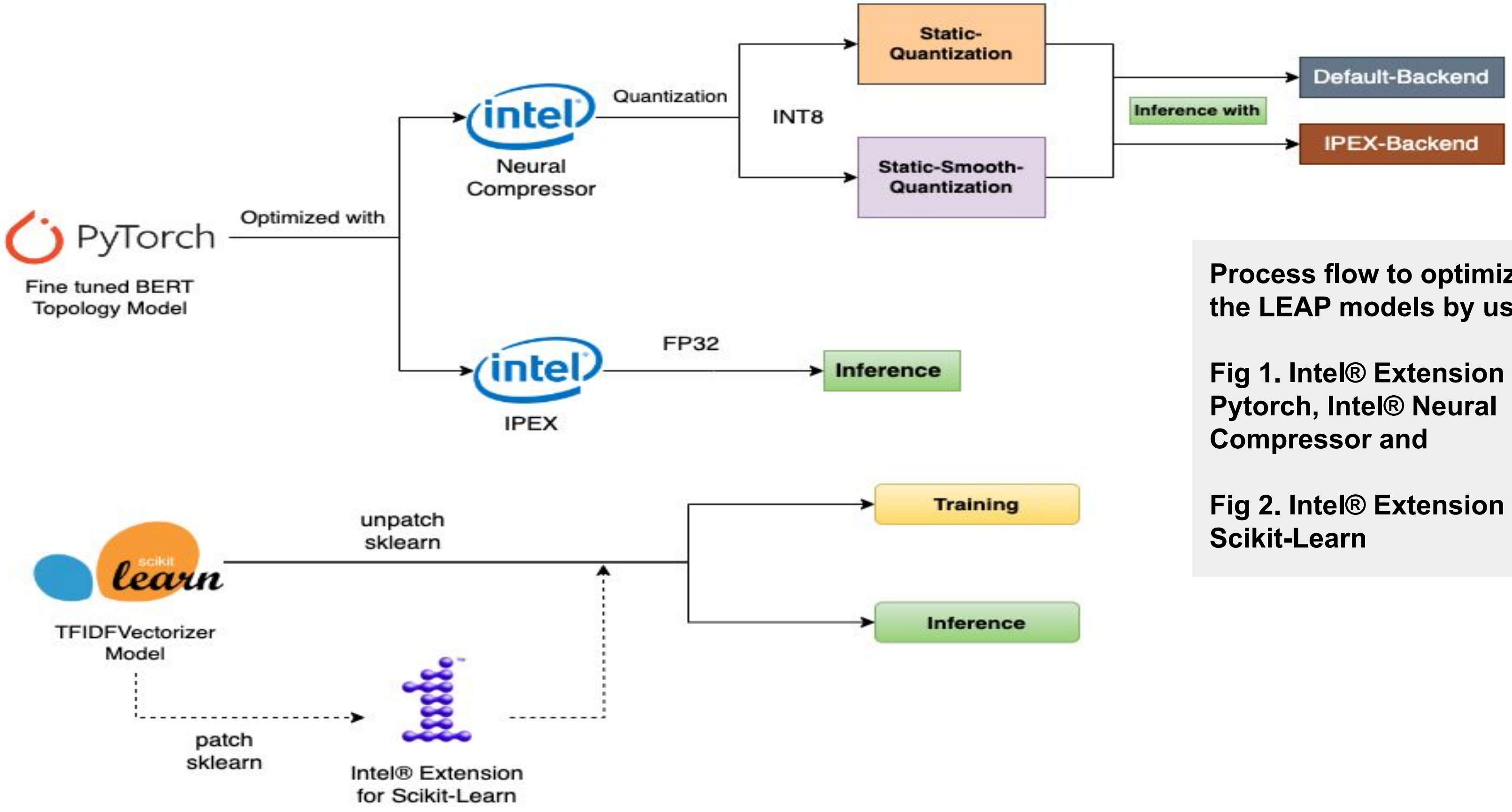


Screen-6 (Interactive Conversational AI Examiner provides hints and motivates a student in case of a wrong answer)

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

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

Result Summary (focus on unique aspects of oneAPI/SYCL that you have used)

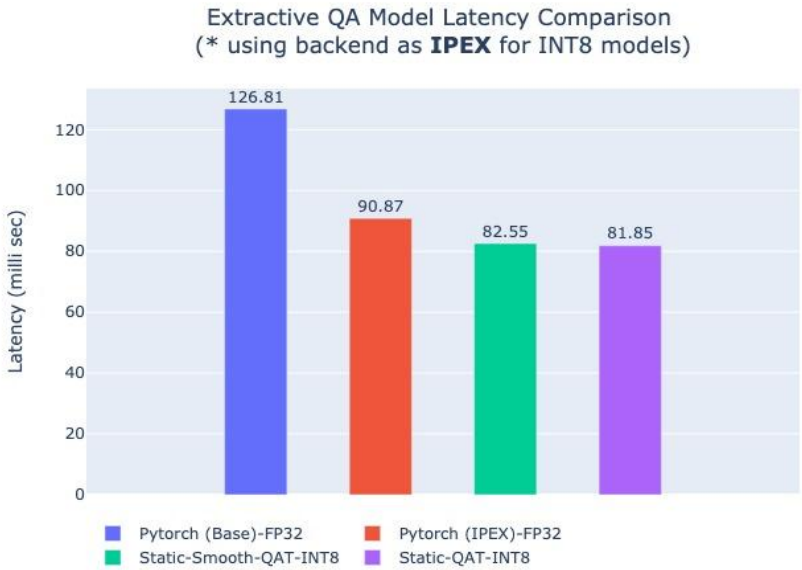


Process flow to optimize the LEAP models by using

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

Fig 2. Intel® Extension for Scikit-Learn

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



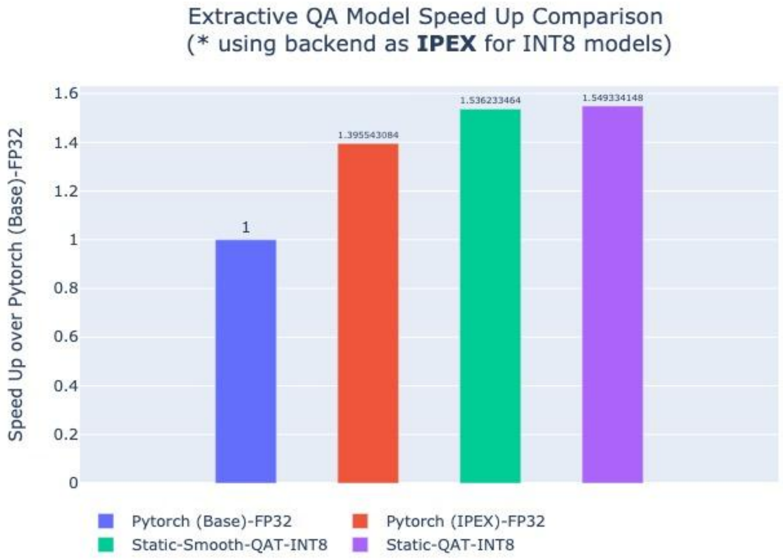
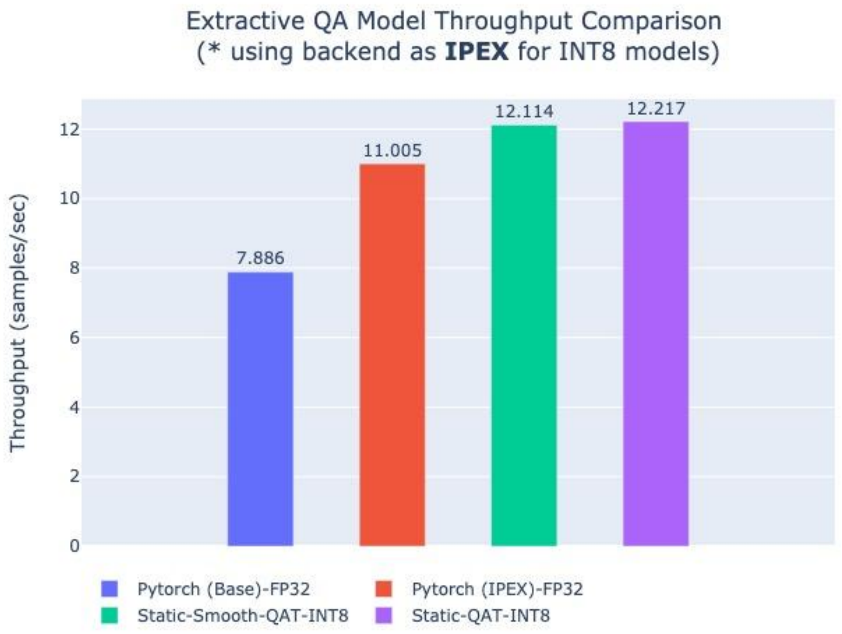


*Fig: Latency/Speed-Up Benchmark result for **our Extractive Question Answering Model (Multilingual)** on Intel® Dev Cloud machine (Intel Xeon Processor (Skylake, IBRS) - 10v CPUs 16GB RAM) with optimization using IPEX-FP32 and Static INT8-Quantization using Intel® Neural Compressor. Please Note that we use backend as **IPEX** for INT8 models here to get further benefit.*

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



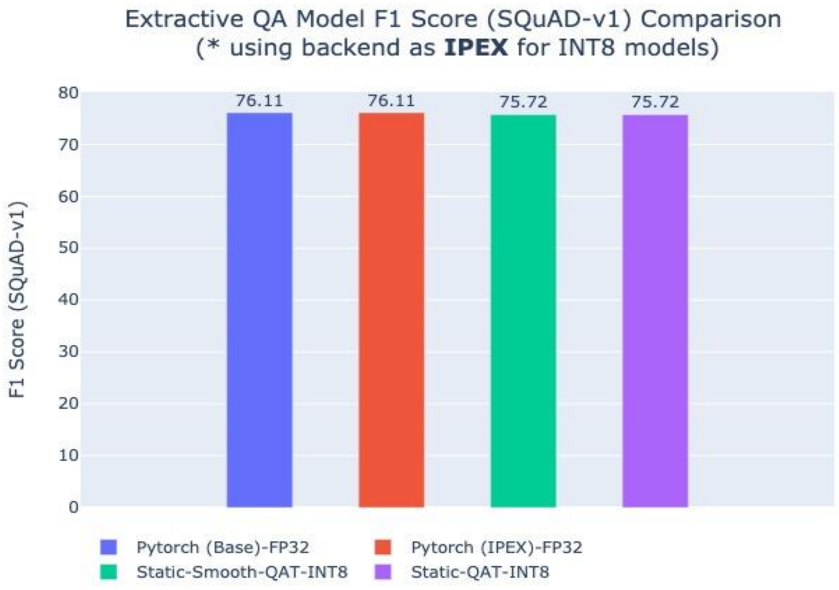
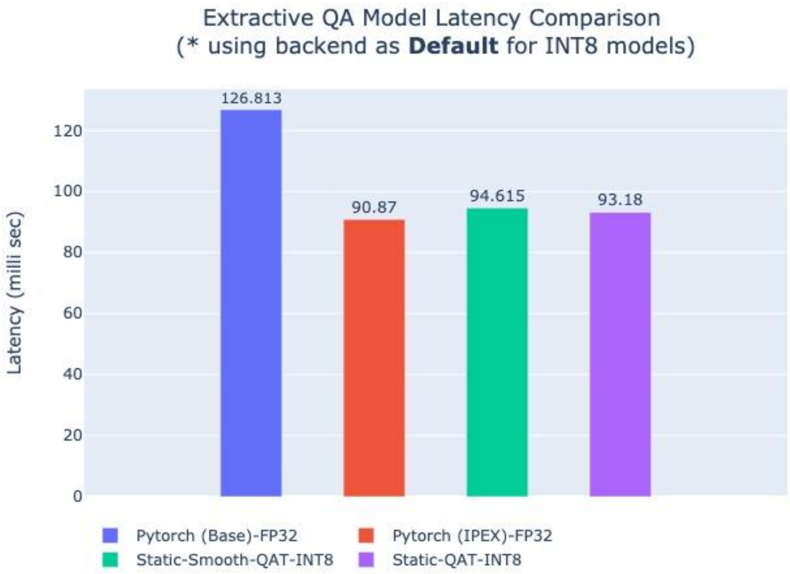


*Fig: Throughput/F1 Score Benchmark result for **our Extractive Question Answering Model (Multilingual)** on Intel® Dev Cloud machine (Intel Xeon Processor (Skylake, IBRS) - 10v CPUs 16GB RAM) with optimization using IPEX-FP32 and Static INT8-Quantization using Intel® Neural Compressor. Please Note that, we use backend as **IPEX** for INT8 models here to get further benefit. Also, the model (https://huggingface.co/ai4bharat/indic-bert) was fine-tuned on SQuAD-v1 dataset.*

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



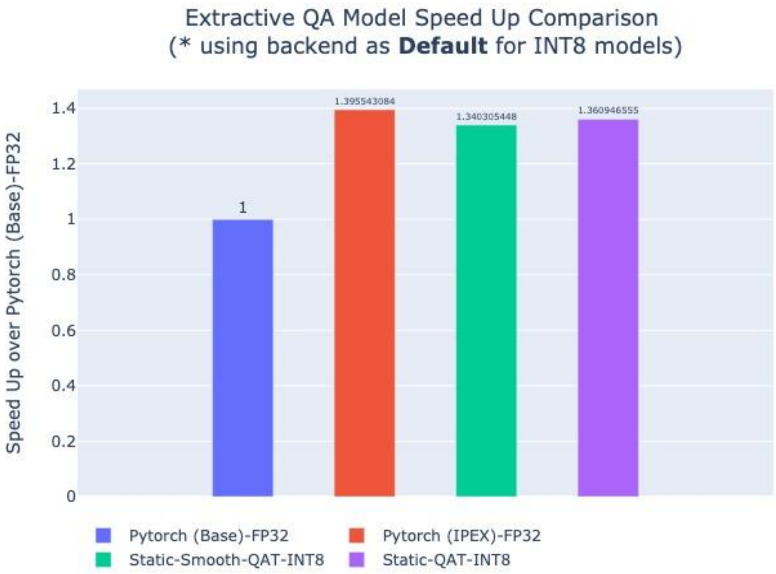
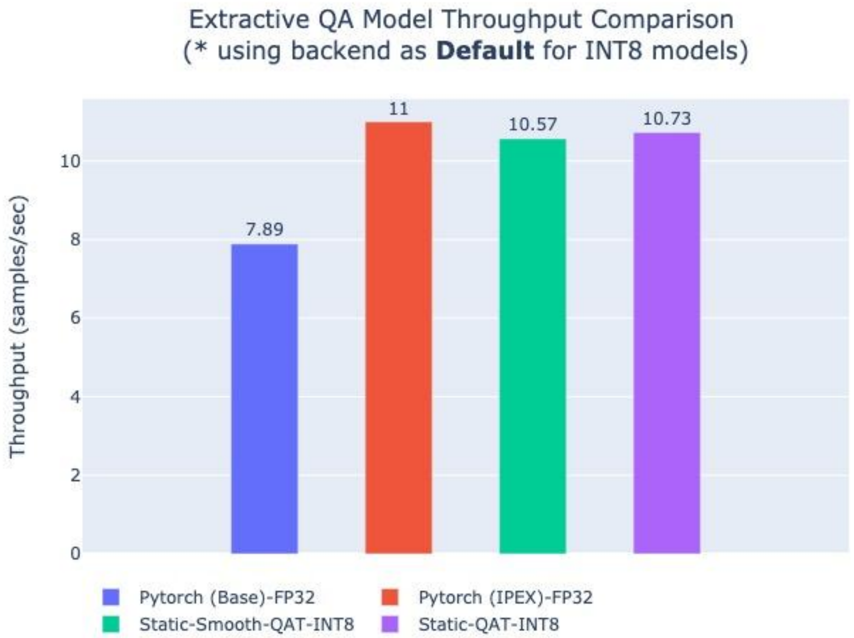


*Fig: Latency/Speed-Up Benchmark result for **our Extractive Question Answering Model (Multilingual)** on Intel® Dev Cloud machine (Intel Xeon Processor (Skylake, IBRS) - 10v CPUs 16GB RAM) with optimization using IPEX-FP32 and Static INT8-Quantization using Intel® Neural Compressor. Please Note that we use backend as **Default** for INT8 models here to get further benefit.*

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



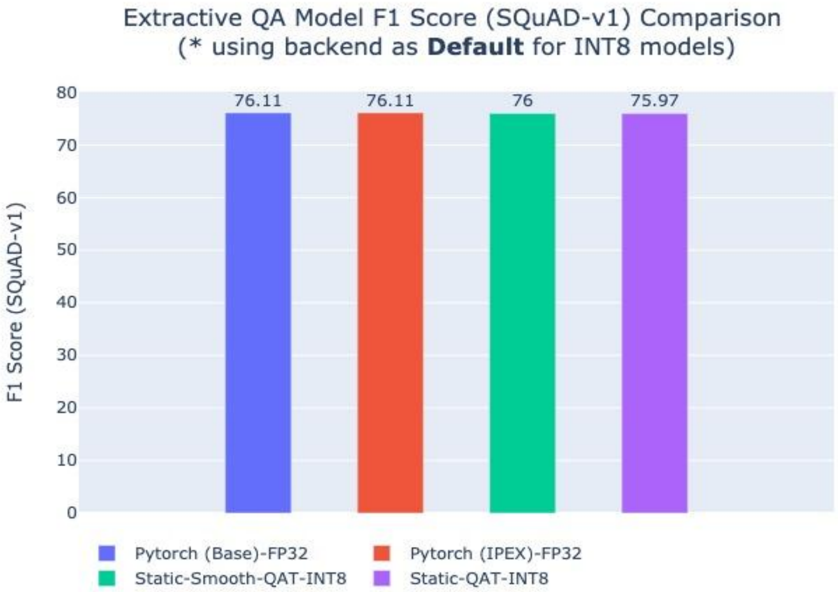


*Fig: Throughput/F1 Score Benchmark result for **our Extractive Question Answering Model (Multilingual)** on Intel® Dev Cloud machine (Intel Xeon Processor (Skylake, IBRS) - 10v CPUs 16GB RAM) with optimization using IPEX-FP32 and Static INT8-Quantization using Intel® Neural Compressor. Please Note that we use backend as **Default** for INT8 models here to get further benefit. Please Note that, the model (https://huggingface.co/ai4bharat/indic-bert) was fine-tuned on SQuAD-v1 dataset.*

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

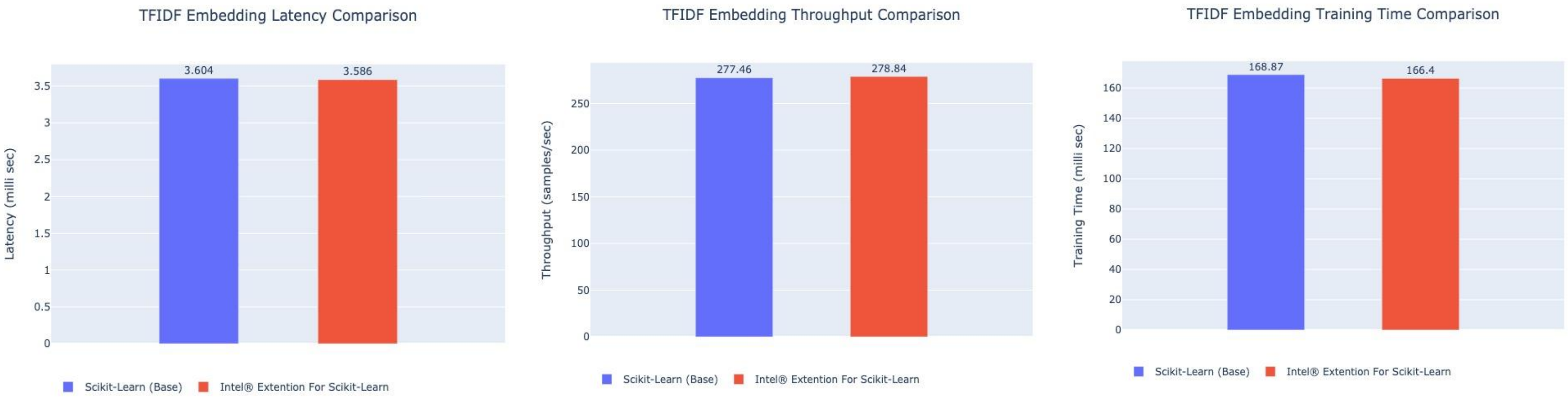


*Fig: Benchmark results for **TFIDFVectorizer** Embedding model during training and inference on Intel® Dev Cloud machine (Intel Xeon Processor (Skylake, IBRS) - 10v CPUs 16GB RAM). Please Note that we don't see much of a difference may be because we used a tiny dataset.*

THANK YOU