Montreal, Canada
+1 (438) 979-2081
saikalyan9981@gmail.com
linkedin.com/in/kalyansid
github.com/saikalyan9981

Sai Kalyan Siddanatham

Education

2021–2023 **University of Montreal (Mila Quebec AI)**, *Masters in Computer Science*, GPA – 4.3/4.3.
Masters in ML with Internship (Co-op), Expected Graduation: Apr 2023

2016–2020 **Indian Institute of Technology, Bombay**, *Bachelors (with Honours) in Computer Science*, GPA – 8.59/10.

Work Experience

May 2022 – Oct 2022 **Research Scientist Intern**, NUANCE-MICROSOFT, Montreal, Canada, *Tensorflow.*
- Optimized **Next Turn Prediction** for a Conversational Agent AI App. Tweaked the Transformer models by **reducing embedding size, ensemble approaches**, resulting in 1% EM improvement with 2X latency reduction.
- Implemented **Spell Checker** for the AI application using **Viterbi Algorithm** with 91% F1, improving baseline by 6%. Created a prototype to streamline the use of the **Bing Spell Checker API** for optimal performance.

Jul 2020 – Jul 2021 **Machine Learning Engineer**, HILABS, Pune, India, *PyTorch, PySpark, Spark, Scala.*
- Built **Provider Network Analysis** App to detect anomalies in the health provider database holding hundreds of thousands of provider contracts and tabular records, using **OCR**, **Bert-NER** and **Pattern mining**.
- Extracted information from searchable and scanned contracts building a pipeline with Tesseract OCR, antiword and Bert-NER. Reduced latency by 10X using **distributed computing with PySpark**.
- Deployed the App on-prem, citing around 40 critical mismatches of provider network and 1000s of minor ones.

May 2019 – Jul 2019 **Research Engineering Intern**, SAMSUNG RESEARCH, Bangalore, India, *PyTorch.*
Explored NLP techniques to **Analyse a conversation, Build context and Answer the questions**. Framed as a question-answering task and finetuned BERT on the SQuAD dataset. Created a WebApp to demo the results.

May 2018 – Jul 2018 **Software Development Intern**, FRANKLIN TEMPLETON, Hyderabad, India, *Django.*
Developed a Web Application using Django for **maintenance of Fixed Income**, helping analysts track decisions

Key Projects

MLOps Specialization **Deploying ML Models in Production**, COURSERA, *Kubernetes, Locust, Istio, TFServing.*
- Deployed ResNet50 and ResNet101 as production and canary models on **Google Kubernetes Engine** using weighted load balancing and request header-based traffic splitting strategies.
- Configured Horizontal Pod Autoscaler and monitored serving performance and resource utilization using locust.
- Automated Machine Learning CI/CD pipeline using **Github Actions** for efficient and consistent delivery.

Data Science **NHL live xG Prediction**, U.MONTREAL, *Sklearn, Plotly, Docker, Pandas, Comet.ml.*
- Developed machine learning models to predict expected goals (xG) of shots in live NHL games using **XGBoost** and logistic regression, with performance measured using AU-ROC and calibration curves.
- Deployed models in a **Docker container** accessible via a REST API and tracked experiments using Comet.ml.
- Performed data acquisition, wrangling, EDA with shot map visualizations, and feature engineering.

Operating Systems **Xv6 Operating System**, IIT BOMBAY, *C++.*
- Used semaphores and mutex locks for **multi-threading** issues such as reader-writer and barrier synchronization.
- Developed a simple **file system** with basic functions and APIs to handle files on an emulated disk interface

Academic Achievements

2022 Excellent grades (A+) in all my graduate courses, including Representation Learning and Graphical Models.

2018 Amongst Top 300 of 4000 teams in ACM-ICPC prelims and Top 150 in Regionals receiving honorable mention

2016 **Rank 42 in JEE-Advanced** across all of India among 1.2 million candidates for admission into IITs

2014 Gold Medalist for being amongst the Top 35 students in Indian National Junior Science Olympiad Camp.

Technical Skills

Languages Python, C++, Scala, Java, SQL, Matlab, Julia

Softwares PyTorch, Tensorflow, Numpy, Pandas, Sklearn, Spark, PySpark, Docker, Kubernetes, Locust, FastAPI, Istio, Comet.ml, Django, Plotly, Flutter, GCP

Coursework NLP, Convex Optimization, Computer Architecture, Software Systems, Automata Theory, Data Structures and Algorithms, Cryptography, Compilers, Computer Networks, Graphical Models, Operating Systems, Image Processing, Speech Recognition, Reinforcement Learning