

VIJAY S KALMATH

<https://vijaykalmath.github.io>

<https://linkedin.com/in/vijay-kalmath>

EDUCATION

Columbia University

New York, USA

Master of Science in Data Science, 3.8/4.0

Aug 2021 - Dec 2022

Coursework: Deep Learning, Advanced Deep Learning, Machine Learning, Representation Learning, Statistical Inference, Algorithms.

B.M.S College of Engineering

Bangalore, India

Bachelor of Engineering in Electronics and Communications, 9.37/10.0

Aug 2014 - May 2018

WORK EXPERIENCE

Lexalytics, Inc.

New York, USA

Natural Language Processing - Machine Learning Engineer

Jan 2023 – Current

- Enhanced multi-lingual key-phrase extraction system with fine-tuned HuggingFace BERT, achieving 0.85 F1 score.
- Optimized BERT model inference pipeline with ONNX runtime accelerator and quantization, reducing API response time by 5%.
- Improved LLAMA2-7B's customer review summary generation with QLORA and LORA techniques with PEFT library.
- Designed GPT-4 prompts to generate 1000+ synthetic reviews for sales presentations, eliminating use of real-world data.

Columbia University, AI Model Share

New York, USA

Graduate Machine Learning Researcher

Aug 2022 – Dec 2022

- Official Maintainer of AIModelshare Python library on conda-forge with 35,576+ downloads.
- Designed and developed model determinism and model replication framework using TensorFlow.
- Engineered pipelines for cloud-based storage and hosting of Keras and PyTorch models, reducing AWS operating costs.

Columbia University, WiMNet Lab

New York, USA

Graduate Deep Learning Researcher

May 2022 - Aug 2022

- Developed models for micro-climate estimation based on noise in radar data, enhancing wind speed accuracy by 30%.
- Increased classification accuracy by 30% with DCT based Fourier features of radar transmission with TensorFlow.
- Developed time distributed CNN with LSTMs for image sequence classification, resulting in 35% increase in F1-Score.
- Performed feature engineering with PCA and TSNE, analyzed 20+ experiment's data to visualize spatial and temporal properties.

Cisco Systems

Bangalore, India

Network Engineer-II-Escalation Engineer

Jul 2018 - Jul 2021

- Diagnosed issues in Cisco ACI's RAFT and Linux Subprocesses on distributed computing systems, for 400+ customers.
- Crafted Data Pipelines for real-time monitoring of Application Centric Infrastructure labs, improved server utilization by 40%.
- Spearheaded 30+ projects to enhance Cisco product usability with containerized log analyzers, low-code and IaC platforms.

PROJECTS

XProNet – Radiology Report Generation for Chest X Ray Images

Sep 2022 - Dec 2022

- Evaluated the real-life applicability of XProNet - a cross-modal prototype learning model for radiology report generation.
- Developed an efficient inference pipeline with a pseudo-label generator, improving CheXBert label accuracy from chest X-Rays.
- Reduced mode-collapse by 15% in generated medical reports through interpretability research of XProNet.

Adversarial Training in Distillation of BERT

Jan 2022 - May 2022

- Investigated impact of teacher-student knowledge distillation on robustness of 3 BERT-like Language Models with PyTorch.
- Trained semi-supervised GANBERT with 50% unlabeled data, performed distillation and evaluated robustness with TextAttack.
- Implemented 4+ gradient-based adversarial data augmentation techniques with GANBERT and DISTILBERT.
- Demonstrated that teacher-student model compression reduced adversarial robustness by 70% during TextFooler Attacks.

Spectral Representations for Convolutional Neural Networks

Sep 2021 - Dec 2021

- Devised customized spectral pooling, frequency dropout, and spectral convolution TensorFlow layers with Fourier transform.
- Attained 80% test accuracy with a 40% decrease of training time with Bayesian Hyperparameter tuned spectral CNN.
- Achieved 2x – 5x times computational speed up with spectral parameterized CNN architectures for ImageNet and CIFAR-100.

SKILLS

Programming Languages – Python, R, JavaScript, C++.

Databases – SQL, Postgres, MongoDB, Elastic-Search.

Python Frameworks – TensorFlow, HuggingFace, PEFT, WandB, PyTorch, Onnx, Scikit-Learn.

Infrastructure Technologies – Spark, Linux, ELK Stack, Docker, Kubernetes, Kafka, AWS, Google Compute, Kubeflow.