Abhishek Singh

📞 858 228 7140 | ✉ abs006@ucsd.edu | 🔗 LinkedIn | asking28 | 🌐 asking28.github.io | Publications

JOB AND RESEARCH EXPERIENCE

Oracle Inc.   Jun 2023 - Present
Machine Learning Software Engineer | MySQL Heatwave AutoML   Redwood City, CA

- Spearheading the development of the MySQL in-database Auto Machine Learning framework to streamline the data science workflow, encompassing various tasks such as Generative AI, Retrieval Augmented Generations (RAG) with Large Language Models (LLMs), Classification, Regression, and more.
- Engaging with customers to understand their needs, providing tailored support through tutorials and demos on production datasets for applications like anomaly detection in DB logs, user ticket classification, and automated policy-based responses.
- Designed end-to-end Data Science pipelines encompassing data cleaning, feature selection, algorithm selection, model tuning, and model explainability techniques.

Graduate Student Researcher   Dec 2021 – May 2023
Prof. Pengtao Xie | Multilevel Optimization in Machine Learning   San Diego, CA

- Developed an innovative solution to address the challenge of fine-tuning Pre-trained Language Models (PLMs) on low-resource datasets by implementing an attention-guided weights mixup strategy.   [NAACL 2024 Paper]
- Leveraged the inter-dependency between task weights and pre-trained weights to formulate the learning objective as a Bi-level Optimization (BLO) problem.
- Designed a novel multilevel optimization algorithm for text augmentation along with its derivation and complexity analysis. The projects goal is to leverage synthetic data and training data to improve the performance of existing translation systems.**(Submitted)**

Brain Technologies Inc.   Jun 2022 – Sep 2022
Natural Language Processing Intern | Machine Learning Team   San Mateo, CA

- Led the implementation of a proof of concept for a location-based recommendation application, integrating NLP techniques for intent recognition, NER, review summarization, and personalized recommendations.
- Implemented T5 model for review summarization from Google Maps API, achieving a **ROUGE score of 25.2**.
- Optimized a BERT-based model for joint intent classification and NER, securing **F1 scores of 93.1** (NER) and **95.6** (intent classification) across 16 classes.

Samsung R&D Bangalore   Jun 2019 – Aug 2021
Machine Learning Research Engineer | Bixby Voice Intelligence Team   Bangalore, India

On-Device System for Device Directed Speech Detection [Paper]

- Pioneered a multi-stage framework distinguishing device-directed utterances, leveraging both speech and textual features, reducing Equal Error Rate (EER) to **3.6%** with Bi-LSTM and attention models.
- Achieved substantial model size reduction for on-device deployment, compressing from **98.5MB and 196MB to 11.3MB and 8.28MB**, respectively.

Joint Intent and Slot Identification from Code Mixed Input [Paper]

- Developed a novel code-mixed dataset for Virtual Assistant tasks and introduced a deep learning model, achieving **96.68% accuracy and 94.3% F1 score**, setting a new standard in the domain.
- Designed a unique architecture combining **Indicative Intent Feature Network (IIFN)** and **Domain Specific Feature Network (DSFN)** for enhanced input understanding.

EDUCATION

University of California San Diego | California, US   Sep 2021 – May 2023
M.S. in Computer Science, Specialization in Machine Learning and Artificial Intelligence   GPA: **3.96 / 4.0**

- **Selected Coursework**: Convex Optimization, Structured Prediction for NLP, Statistical NLP, Deep Generative Models, Machine Learning Theory, Unsupervised Learning, Data Systems for ML, Causal Inference, and more.

Indian Institute of Technology (Banaras Hindu University) | Varanasi, India   Jul 2015 – May 2019
B.Tech. in Electrical Engineering   GPA: **9.09 / 10**

- **Selected Coursework**: Machine Learning, Data Mining, Probability & Statistics, Calculus, Ubiquitous Computing.

SKILLS

Programming Languages : Python, C++, SQL, JAX
Frameworks/Tools : CVXPY, GPyOpt, PyTorch, TensorFlow, Pandas, Numpy, Scikit learn, Spacy, Spark NLP, PyMC, Optax
MLOps/Dev Tools : Kubernetes, AWS, Docker, Git, LaTeX, VS Code, Visual Studio, Jupyter Notebook, Weights and Biases

PROJECTS

Image segmentation on Synthetic Aperture Radar at various locations: [Github] Dec 2022 – Mar 2023

- The images are 256 x 256 pixels and each pixel is classified as non-seep (0) or 7 classes of seeps (1-7). The objective of the exercise is to segment regions that contain seeps, and as an optional task to classify the seeps. Data is highly imbalanced across different classes.

- Implemented state-of-the-art image segmentation $UNet++$ model with resnet encoder on Pytorch. Trained the model using Tversky Loss and also monitored DICE and $f1$ while training. The model achieves an IOU of 0.9486 and F1 of 0.9736.

- Implemented state-of-the-art image segmentation $UNet++$ model with resnet encoder on Pytorch. Trained the model using Tversky Loss and also monitored DICE and $f1$ while training. The model achieves an IOU of 0.9486 and F1 of 0.9736.

Hypothesis Testing on Cookies Cats game for player retention December 2022 – Jan 2023

- In this project, we analyzed the result of an A/B test where the first gate in Cookie Cats was moved from level 30 to level 40. We checked statistical significance difference between two groups to determine player retention.

- Since Shapiro Testing rejected the H0 hypothesis of normality assumption, we had to utilize a non-parametric test, specifically the Mann Whitney U test, to compare the two groups. Subsequently, the Mann Whitney U test rejected the H0 hypothesis, leading us to conclude that the A/B groups are dissimilar.

Reinforment learning algorithms on different games Jan 2022 – Mar 2022

- Designed tic-tac-toe as Markov Decision Process (states, action, and rewards). Implemented value iteration algorithm and trained until convergence of each state.

- Implemented Monte Carlo policy evaluation, Temporal-difference policy evaluation, and Q-learning for playing Blackjack.

- Trained expectimax algorithm with added heuristics to play 2048 on a python-based simulator. Max score achieved was **35172**.

Variational Auto Encoder and GAN models for chest x-ray generation: [Report-1] [Report-2] Jan 2022 – Mar 2022

- Trained and compared Variational autoencoder and Conditional Generative Adversarial Network based models for Chest X-ray image generation.

- The quality of generated images is tested using Inception Score and Frechet Inception Distance (FID). CGAN achieves **Inception Score of 1.64 and FID of 268** whereas, VAE achieves **Inception Score of 1.45 and FID of 276.98.**

Joint abstractive and extractive method for long document summarization: [Paper] [Code] Sep 2020 – Mar 2021

- Proposed joint model of Pointer Network and T-5 (Text-to-text transfer Transformer) algorithms jointly trained using Policy Gradient Reinforcement Learning algorithm by optimizing BLEU score.

- Method is evaluated on ROUGE-N (1,2), LCS,and SU4. It achieves **ROUGE-LCS** F1-score of **45.6**, surpassing other methods.

Method for Sentiment and Offensiveness detection in Social Media: [Paper] [Code] Mar 2020 – Sep 2020

- Proposed stacked ensemble-based deep learning model for five languages, including two code-mixed languages.

- Model achieved 0.886 F1-Macro on offensiveness detection in the Greek language, 0.9 in English Language, and an F1 score of 0.756. Overall system stood third in the competition.

Academic Experience

Teaching Assistant

- Convex Optimization CSE 203B, Statistical Natural Language Processing CSE 156, Graduate Linear Algebra Course ECE 269, Computer Programming and Linux Concepts, Basic Electrical Engineering

Reviewer

- Reviewer for ICML 2022, EMNLP 2022 (Industry Track).