

Abhishek Singh

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

EDUCATION

University of California San Diego | California, US

Sep 2021 – May 2023

M.S. in Computer Science (*Specialization: Machine Learning and Artificial Intelligence*)

GPA: **3.96 / 4.0**

- **Coursework:** Convex Optimization, Structured Prediction for Natural Language Processing, Search & Optimization, Probabilistic Reasoning, Statistical NLP, Linear Algebra and Applications, Unsupervised Learning, Deep Generative Models, AI Seminars, Interpretability and Explainability, Data Systems for Machine Learning, Machine Learning Theory, Causal Inference

Indian Institute of Technology (Banaras Hindu University) | Varanasi, India

Jul 2015 – May 2019

B.Tech. in Electrical Engineering

GPA: **9.09 / 10**

- **Coursework:** Machine Learning, Data Mining, Mathematical Modeling, 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

JOB AND RESEARCH EXPERIENCE

Oracle Inc.

Jun 2023 - Present

Machine Learning Software Engineer II | MySQL Heatwave AutoML

Redwood City, CA

- * Developing Auto Machine learning framework for different data modalities.
- * Framework has optimized end-to-end pipeline from feature engineering, model training to inference.

Graduate Student Researcher

Dec 2021 – May 2023

Prof. Pengtao Xie | Multilevel Optimization in Machine Learning

San Diego, CA

- * **AutoML:** Designed a **novel multilevel optimization** algorithm for text augmentation along with its derivation and complexity analysis. The project's goal is to leverage synthetic data and training data to improve the performance of existing systems.
- * Trained the framework on Machine Translation Multi30k dataset. The proposed framework performs better than existing augmentation techniques on the held-out test set. It shows improvement of **2.1 BLEU, 2.5 METEOR, and 1.6 ROUGE scores**.
- * The same framework, when trained on text summarization dataset XSum, shows improvements of **1.8 BLEU, 1.3 METEOR, and 1.5 ROUGE scores. (Paper under review at Transactions of ACL)**
- * **Fine-Tuning:** Designed **novel bilevel** based framework that aims to fine tune pretrained models using a combination of pretrained and finetuned weights. Every weight is associated with their contribution factor which is optimized during training. (In progress)

Brain Technologies Inc.

Jun 2022 – Sep 2022

Natural Language Processing Intern | Machine Learning Team

San Mateo, CA

- * Developed end-to-end proof of concept project for road trip planning. Involves multiple NLP applications such as intent recognition, NER extraction, review summarization, and location recommendations.
- * Used reviews extracted from google maps API to train transformer architecture-based T5 model for the summary generation with **ROUGE score of 25.2** on the test set.
- * Trained joint intent classification and Named Entity Recognition model using transformer based BERT model. Achieved an **F1 score of 93.1 on NER task and 95.6 on Intent Classification task** with 16 classes.
- * Used GPT3 model for few-shot prompt based learning to develop users' taste profile for food recommendation system.

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 for Improving Human-Computer Interaction [Paper]

- * Developed multi-stage framework to classify utterances intended towards device using speech and textual features.
- * Evaluated the framework using Equal Error Rate(EER) to find a balance between false acceptance and rejection. Stage-1 Bi-LSTM-based model utilizes speech signal features such as MFCC, rate of speech, STFT, DCT, and text-based features. Achieved **3.6% of EER** on optimized model.
- * Designed stage-2 model to process contextual input. Proposed Natural Language Inference model using the attention-based model. Achieved an **EER of 5.1%** on Stage-2 model.
- * Optimized models' sizes for on-device functionality from **98.5MBs to 11.3MBs and 196MBs to 8.28MBs**.

Joint intent and slot identification from Code Mixed Input for Virtual Assistants [Paper]

- * Introduced first of its kind, a code-mixed dataset covering 10 most preferred features for a Virtual Assistant, and developed task-specific deep learning based model.
- * Conducted empirical analysis by comparing the suitability & performance with various state-of-the-art methods. The proposed method attains **96.68% accuracy and 94.3% F1** on the dataset.
- * Proposed architecture includes two major parts: a character-based network, **Indicative Intent Feature Network (IIFN)** and **Domain Specific Feature Network (DSFN)**.

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.

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 H_0 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 H_0 hypothesis, leading us to conclude that the A/B groups are dissimilar.

Reinforcement 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).