mt1180770@iitd.ac.in
subhalingam.d@gmail.com

Subhalingam D

linkedin/subhalingamd
github/subhalingamd
subhalingamd.me

Education

Bachelor of Technology in Mathematics and Computing, Indian Institute of Technology, Delhi, *CGPA: 8.15/10* **2022**
All India Senior School Certificate Examination, Chennai Public School, Chennai, *Percentage: 96.4%* **2018**
All India Secondary School Examination, Chennai Public School, Chennai, *CGPA: 10/10* **2016**

Experience

KnowDis Data Science LLP, Delhi **Jan 2022 – Present**
Data Science Intern: *Product Search and Recommendation Engine (for IndiaMART)* MARS Team
- Generated paraphrased queries from fine-tuned *doc2query-T5* and trained an auxiliary multiclass classification model using *GPT-2* for **product tagging**
- Using the auxiliary model to decrease latency of *mBART*-based Generative Entity Retrieval (GENRE) model by **rescoring the top-k retrieved candidates**

Samsung R&D Institute, Delhi **Jun 2021 – Jul 2021**
Software Engineering Intern: *Sound Source Localization, Tracking and Separation* Intelligence Software Team
- Developed **sound source direction estimation** module, using time delay of arrival between pairs of microphones in an array, for real-time audio input
- Added modules for mapping & tracking active sources and extracting signals of individual sources for downstream object identification pipeline
- Integrated stationary noise estimation module for ambient noise removal and reduced maximum direction of arrival error to 7°

MateRate Education Pvt. Ltd., Delhi **May 2020 – Jul 2020**
Machine Learning R&D and Backend Web Development Intern: *Students' Latent Knowledge Space Modelling and Results Portal Development*
- Developed Item Response Theory-based probabilistic models to estimate and analyze the ability of 5000+ students & difficulty of 200+ questions
- Designed database schema & workflows, devised **system design** for production, built Web APIs using **Django REST framework** and tested them
- Deployed *Django*-based backend to **EC2** with *MySQL* database layer using **RDS**, and *React*-based frontend to **S3** with **CloudFront** CDN integration
- Set up **Auto Scaling group** with scaling policies (triggered by **CloudWatch** alarms) and attached **Load Balancer** for traffic-based horizontal scaling

Skills

Development: **Python**, **Java**, C++, **C**, Bash, **Django**, **FastAPI**, **AWS**, SQL, PHP, MATLAB, HPC Cluster, Docker, Git, CSS, jQuery
NLP/ML Tools: **PyTorch**, Transformers, PyTorch-Lightning, TensorFlow, Keras, NLTK, Selenium

Projects

Tracking State Changes for Entities in Technical Procedural Text **Research Project, Data Group,** Feb 2021 – Present
Prof. Srikanta Bedathur & Prof. Maya Ramanath | In collaboration with **IBM AI Horizons Network** [arXiv preprint]
- Prepared a dataset by **scraping WikiHow pages** from *Computers and Electronics* category and developed **knowledge-injected BERT-based baselines**
- Deployed NLP scheme to predict the actions taken & changes in properties of entities involved at each step of a process and performed error analysis
- Currently, surveying the literature and building models to predict the **next most-likely action** to be performed given a sequence of performed actions

chaii - Hindi and Tamil Question Answering **Prof. Mausam, NLP Course,** Oct 2021 – Nov 2021
- Fine-tuned *XLM-RoBERTa* for **multilingual Q/A** using *chaii-1* dataset augmented with *MLQA*, *XQuAD* & *SQuAD* and attained test **Jaccard score of 68.72%**

Rule-based Written-to-Spoken Text Converter **Prof. Mausam, NLP Course,** Aug 2021 – Sep 2021
- Built a **regex**-based system that accounts for chunks with abbreviations, dates. numerical quantities and inflections; obtained test **F1-score of 97.94%**

Corporate Bankruptcy Prediction **Prof. Niladri Chatterjee, Data Mining Course,** Feb 2021 – Apr 2021
- Reviewed state-of-the-art bankruptcy prediction models & observed poor recall; hypothesized **class imbalance** & **missing values** to be the reasons
- Trained an ensemble model with **Mean Imputation** & **SMOTE** transformations on *Polish companies* dataset and **gained 10% improvement in recall**

Extended Vector Space Model for News Articles Retrieval **Prof. Srikanta Bedathur, IR Course,** Oct 2020 – Nov 2020
- Implemented an end-to-end retrieval system indexed using **TF-IDF weights**, with support for prefix search & named-entity based filters at query time
- Reduced index size by half with **gap encoding**; applied **pseudo-relevance feedback** based probabilistic **query expansion** to rerank retrieved articles

More projects:
- **Bachelor's Thesis: Identification of Hate Spreaders on Social Media**: Extracting features from data, analyzing them and building baselines *(Ongoing)*
- **Context-Sensitive Word Sense Disambiguation**: Compared non-contextual and contextual embeddings (GloVe+BiLSTM vs BERT) using WiC dataset
- **Tweet Sentiment Classifier**: Processed tweets with tweet normalization, internet slang dictionary, stemming, etc.; vectorized with TF-IDF; fed into LR
- **Adaptive Neuro-Fuzzy Inference System for Diabetes Prediction**: Trained a *Takagi–Sugeno* type fuzzy neural network with a test accuracy of 81.3%

Activities & Positions Held

Teaching Assistant, Information Retrieval and Web Search, *Prof. Srikanta Bedathur* Aug 2021 – Dec 2021
General Secretary, Mathematics Society, IIT Delhi Aug 2021 – Present
Web Development Executive, Entrepreneurship Development Cell, IIT Delhi Sep 2019 – Jun 2020
Volunteer in Teaching projects, National Service Scheme (NSS), IIT Delhi Dec 2018 – Present