Subhalingam D

Data Scientist / Software Engineer

mt1180770@maths.iitd.ac.in
contact@subhalingamd.me

subhalingamd.me
github/subhalingamd
linkedin/subhalingamd

EDUCATION

Bachelor of Technology in Mathematics and Computing, *Indian Institute of Technology, Delhi, CGPA: 8.196/10*   May 2022
All India Senior School Certificate Examination, *Chennai Public School, Chennai, Percentage: 96.4%*   May 2018
All India Secondary School Examination, *Chennai Public School, Chennai, CGPA: 10/10*   May 2016

EXPERIENCE

KnowDis Data Science LLP, Delhi   **May 2022 – Present**
Data Scientist

- Product Search and Recommendation Engine (for e-commerce)
- English-Hindi Translator with Style Transfer (for open domain)
- Indic Languages Translator (for e-commerce and open domain)

KnowDis Data Science LLP, Delhi   **Jan 2022 – May 2022**
Data Science Intern | *Product Search and Recommendation Engine (for IndiaMART)*   *MARS Team*

- Devised a **GPT-2**-based classification model and added heuristics based on analysis for tagging product categories from **97k labels**
- Fine-tuned **T5** model for query paraphrasing & queries generation from product categories to deal with tail categories in the data
- Outperformed previous seq2seq system (**3x faster response time**; +1% accuracy) and was **deployed to production** on IndiaMART

Samsung R&D Institute, Delhi   **Jun 2021 – Jul 2021**
Software Engineering Intern | *Acoustic Sound Source Localization, Tracking and Separation*   *Intelligence Software Team*

- Developed **sound source direction estimation** module using time delay of arrival of signals between pairs of microphones in array
- Added modules for tracking active sound sources and extracting individual signals for downstream object identification pipeline
- Integrated stationary noise estimation module for ambient noise removal and reduced maximum direction of arrival error to 7°
- *Received **Pre-Placement Offer (PPO)** for impeccable performance during internship*

MateRate Education Pvt. Ltd., Delhi   **May 2020 – Jul 2020**
Machine Learning and Web Development Intern | *Students' Latent Knowledge Space Modelling and Results Portal Development*

- Developed Item Response Theory-based models to estimate and analyze the ability of 5000+ students & difficulty of 200+ questions
- Designed database schema, devised system design for production, built Web APIs using **Django REST framework** and tested them
- Deployed *Django* backend to **EC2** with *MySQL* database on **RDS** and *React* frontend to **S3** with **CloudFront** CDN integration
- Set up **Auto Scaling group** with scaling policies (triggered by CloudWatch alarms) and attached **Load Balancer** for horizontal scaling

SKILLS

Development: **Python**, **Java**, C++, C, Bash, **FastAPI**, **Django**, **AWS**, SQL, MATLAB, HPC Cluster, Docker, Git, CSS, jQuery, Streamlit
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 – Apr 2022
Prof. Srikanta Bedathur & Prof. Maya Ramanath | *In collaboration with **IBM AI Horizons Network***   *[arXiv preprint]*

- Prepared a dataset consisting of *How-to* troubleshooting FAQs by **scraping WikiHow pages** from *Computers and Electronics* category
- Constructed knowledge-injected **BERT**-based baselines to **predict changes in properties of entities** involved at each step of a process
- Surveyed the literature to build **next-step recommender** from given sequence of performed actions and developed LSTM baselines

Identification of Hate Spreaders on Social Media   Prof. Niladri Chatterjee, Bachelor's Thesis, Jan 2022 – Apr 2022

- Developed baselines with extracted features to identify hate speech spreaders and observed high feature importance for sentiment scores
- Proposed a novel scheme that uses **GloVe** embeddings for encoding and **sentiment scores** as weights to mark word importances
- Evaluated on *PAN@CLEF 2021* dataset and attained an **accuracy of 76%** for English language *(+1% than best)* & 77% with an ensemble

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 + MLQA + XQuAD + SQuAD* datasets and attained Jaccard score of 68.72%

Rule-based Written-to-Spoken Text Converter   Prof. Mausam, NLP Course, Aug 2021 – Sep 2021

- Built **regex**-based system to identify & convert abbreviations, dates/times & numerical quantities to spoken form with **97.94% F1-score**

Corporate Bankruptcy Prediction   Prof. Niladri Chatterjee, Data Mining Course, Feb 2021 – Apr 2021

- Inspected bankruptcy prediction models and 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% recall**

Subhalingam D

mt1180770@maths.iitd.ac.in
contact@subhalingamd.me

subhalingamd.me
github/subhalingamd
linkedin/subhalingamd

Data Scientist / Software Engineer

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

More projects:
- **Context-Sensitive Word Sense Disambiguation**: Studied disambiguation capability of *BERT* and *GloVe+BiLSTM* using *WiC* dataset
- **Tweet Sentiment Classifier**: Vectorized tweets using *TF-IDF* after pre-processing and fed into an *LR* classifier; attained 78.33% accuracy
- **Adaptive Neuro-Fuzzy Inference System for Diabetes Prediction**: Trained a *Takagi–Sugeno* type system with an accuracy of 81.3%

ACTIVITIES

General Secretary, Mathematics Society, IIT Delhi   Aug 2021 – Jul 2022
Teaching Assistant, Information Retrieval and Web Search, *Graduate course offered by Prof. Srikanta Bedathur*   Aug 2021 – Dec 2021
Web Development Executive, Entrepreneurship Development Cell, IIT Delhi   Sep 2019 – Jun 2020
Volunteer in Teaching projects, National Service Scheme (NSS), IIT Delhi