Sumukh S

ssumukh.github.io ✉ s.sumukh11@gmail.com github: ssumukh LinkedIn GScholar +91 94824 19972

EDUCATION

International Institute of Information Technology, Hyderabad (IIIT-H) — Hyderabad, India
B Tech (with Honors) + MS in Computer Science — *(Expected) June 2021*
Integrated dual degree program, with specialization in Machine Learning and NLP
Thesis advisor: Professor Manish Shrivastava, Natural Language Processing(NLP) lab, LTRC

WORK EXPERIENCE

LinkedIn — Bangalore, India
Data Scientist Intern — *May. 2020 – Aug. 2020*

- Flagship Data Science team: To start recommending events through *My Network tab* from the time of creation, I developed models using unsupervised ML and analysed the results. Experimented with GSDMM & LDA + word embeddings.
- Drove a *14% lift in invite acceptance rate*. Preliminary analysis with LinkedIn's *Interest Graph* was also done for the same.
- Worked on calculating CTR for event LIVE Video notifications on a daily basis to update the database & analysed the results. In the process, I identified a discrepancy in the field of a database that was used in a key metric calculation.
- *Used: Python, scikit-learn, gensim, Hive/SQL*

Indian School of Business (ISB) — Hyderabad, India
Software Engineering (ML) Intern — *Aug. 2019 – Nov. 2019*

- To automate the loan approval process and enable banks to analyse small transactions, I built MVPs (standalone webapp & API) for the pilot tests using Django & developed predictive models using Statistical Machine Learning (ML) algorithms.
- After careful feature selection, achieved Recall of 0.64 & F1 score 0.73 for loan automation process.
- The results from the pilot test for the loans using the MVP that I built led to a journal publication. **(Publication Link)**
- *Used: Python, scikit-learn, XGBoost, Django, MySQL/SQL*

Onward Assist — Hyderabad, India
Data Science (Research) Intern — *May. 2019 – June. 2019*

- Worked on identification of cancer in liver tissue WSIs & segmentation of affected areas using Deep Learning algorithms for Computer Vision. Used U-Net FCN architecture to achieve a Jaccard Index of 0.654 and 0.647 for the tasks.
- *Used: Python, PyTorch, scikit-learn*

MarketFront Software Solutions — Hyderabad, India
Software Engineering Intern — *Aug. 2017 – Nov. 2017*

- To help users fetch and update details of their products and the track their inventory across different platforms, I led the group of 3 to build an android application, a webapp and a RESTful API. Custom/new QR codes could also be generated.
- *Used: Java [Android app development], Django, MySQL, Python, ReactJS/Javascript, HTML5, CSS3*

IIIT Hyderabad — Hyderabad, India
Teaching Assistant — *Spring 2019, Monsoon 2020*

- Advanced NLP (*Graduate level course - covers Deep learning + NLP*) *in S19*, Data & Applications *in M20* (*200+ students*).

RESEARCH EXPERIENCE & PUBLICATIONS

Natural Language Processing (NLP) Lab, IIIT Hyderabad — Hyderabad, India
Research Assistant — *May 2018 - Present*

- ***Publication:*** Detection and Annotation of Events in Kannada at LREC 2020 (Workshop), Marseille, France (**Paper Link**)
- Under the guidance of Professor Manish Shrivastava, working on information extraction from unstructured data.
- Previously: Multimodal representation for Visual Question Answering & Machine Translation for low resource languages.

PROJECTS (selected)

Search Engine for Wikipedia — Python (link)

- Used BSBI to create an inverted index of the entire WikiPedia dump to query on & retrieve top results(relevance ranking)

Custom Language Compiler — C++, STL (link)

- Developed a fully functional front-end of the compiler to generate LLVM IR for a custom C-like programming language

HTTPS Multithreaded Proxy Server with Cache — Python (link)

- Implemented a proxy server that can handle multiple clients at a time (multithreading) with server-side caching.

Mini-Dropbox: P2P File Sync — C++ (link)

- Application Level program to keep two local/remote directories synced, similar to Dropbox using sockets.

SQL Engine — Python (link)

- Built a small SQL engine with support for basic queries, joins and aggregate functions with exhaustive error handling.

TECHNICAL SKILLS & RELEVANT COURSEWORK

Programming Languages: Python, C, C++, Javascript, SQL, MATLAB, Java, Bash
Libraries and Tools: PyTorch, scikit-learn, NLTK, Pandas, OpenGL, Android Studio, Git, Flask, HTML, LaTeX, Linux♡
Relevant Coursework: Software Engineering, Database Systems, Computer Networks, Compilers, Distributed Systems, Algorithms, Data Structures, Intro to ML, Natural Language Processing(1 & 2), Optimization Methods, Multivariate Analysis