Nishan Pantha

+1 9389007469 | nishanpantha@gmail.com | linkedin.com/in/nishparadox | github.com/NISH1001

Experience

Graduate Research Assistant Jan 2022 – Present

NASA Interagency Implementation and Advanced Concepts Team (IMPACT), UAH

- Built an extensible framework for benchmarking large-scale data transfer (**5 PetaBytes**) from ESA to NASA. GitHub Repository
- Collaborated with NASA's Bio-Phyiscal Science team and developed an ML-based gene ranking algorithm for gene expressions from space-flown rodents and identified a few important genes (out of **25k** genes) that affect the physical attributes (*gender*, *age*,...) of the rodents. Presented this work as a poster for AGU 22 and waiting for the final paper to be published. Final Slide Deck, Poster Link
- Currently building an extensible evaluation framework for Large Language Models on downstream Earth-Science tasks.

Senior Data/Research Scientist & ML Engineer Nov 2019 – Dec 2021

Docsumo *Kathmandu, Nepal*

- Developed different internal tooling to scale up Docsumo's ML infrastructure.
- Built an end-to-end document table extraction pipeline *(ensemble of model such as CascadeTabNet, DBScan clustering, table header detection, pattern-matching, etc.)* with **row extraction accuracy of 85%+**.
- Built a model-caching framework that helped **reduce the extraction latency by half (30s on average)** on any downstream pipeline.
- Worked on transformer-based Language Model *(modified-BERT with 2d position embedding)* for key-value extraction that significantly **increased accuracy to 90%+** on documents like *Invoices*, *Receipts*, etc.

Co-Founder, Directory of Technology May 2018 – April 2019

MPercept Technology *Kathmandu, Nepal*

- Responsible for all the technical decisions for all the projects such as *real-time face recognition*, *vehicle damage segmentation*, etc.
- Helped co-found an AI community in Nepal, *AIDevNepal*, where we organized 14 Saturdays workshops teaching and mentoring **400+ students** on Data Science and Machine Learning.
- Went to Amman, Jordan as a big-data consultant to **Umniah Telecommunication** to help migrate their RDBMS data to HDFS in **2 months**. During this time, I also worked on real-time CDR file migration using streamsets.

Software Engineer Nov 2016 – Feb 2018

Fusemachines *Kathmandu, Nepal*

- Worked on research projects related to chatbot and Intelligent Character Recognition (ICR). We were able to achieve ICR accuracy of **70%** using RNNs.
- Built recommendation engine as a POC for ML systems.
- Built Salesforce integration API for a client *(Enhatch)*

Data Science Intern Jun 2016 – Aug 2016

Phunka Technologies *Kathmandu, Nepal*

- Worked on data crawling and analytics system using pandas, scrapy, Django, etc.

Miscellaneous 2016– 2021

Miscellaneous *Nepal*

- From *May 2019-July 2019* taught *Artificial Intelligence* course for the final year B.E. Computer Engineering students **(20)** as a part-time lecturer at *Janakpur Engineering College, Kathmandu*. Course slides
- From *August 2018-Feb 2019*, I taught 2 batches of students **(10 per each)** as a course instructor for Data Science at *MPercept Academy*.

Education

The University of Alabama in Huntsville (UAH) Huntsville, AL

MS in Computer Science *Jan 2022 – Present (**Expected Graduation Summer 23**)*

Tribhuvan University, Institute of Engineering, IOE, Pulchowk Campus Kathmandu, Nepal

Bachelor of Engineering in Computer Engineering *2012 – 2016*

Projects

Damage Segmentation Detection | *Python, PyTorch, Tensorflow, CNN*

- This is a project collaboration between MPercept Technology and a Germany-based automobile insurance company to help identify damage segments from Vehicle images which are then used for calculating insurance costs.
- Worked on image processing pipeline and built Deep Learning models such as Mask-RCNN, and U-Net to detect damaged segments. We were able to achieve **mAP score of 74%**.

humT: Query by humming system, BE final year project | *Python, numpy, signal processing*

- Worked on a temporal pattern-matching algorithm like Dynamic Time Warping(DTW). Using DTW and audio segmentation, we were able to **correctly identify 30/35 songs** with **top-5 accuracy of 90%**.
- Published at ICTAES, GitHub repo

anuwadak: Statistical Nepali-English text translation | *Python, numpy, Markov Models*

- As a BE third-year project, I Worked on Markov and N-gram models for predictive text generation. GitHub repo

playx: open-source music assistant for Linux | *Python, numpy, beautifulsoup*

- This project has nurtured **more than 200 stars** where I worked on cores such as crawler for songs+lyrics, song-caching mechanism, and string matching algorithm.
- Developed Markov-model-based song recommendation algorithm that uses user logs to auto-generate playlists.
- Experimented on item2vec model to embed songs and auto-generate playlists.
- GitHub repo

tag-generator: generate tags from texts | *Python, numpy, TF-IDF, NLP*

- This is an open-source project to generate *relevant* tags from given texts/documents using TF-IDF, which has gathered **more than 45 stars**. GitHub Repo

panim: open-source mathematical animation tool | *Python, numpy, matplotlib*

- My open-source tool to implement (from scratch) various mathematical animations such as L-Systems, fractals, simulations, etc. GitHub Repo

Technical Skills

Languages: Python(8+ years), C/C++(2+ years), Java, Bash, scala
Database + Tools: SQL, postgres, mongoDB, Docker, kubernetes
Data Science + ML Libraries: numpy, scikit-learn, scipy, scikit-optimize, mlflow, pandas, matplotlib, spacy
Deep Learning Frameworks: PyTorch, Tensorflow, transformers, sentence-transformers, mmcv
Web Technologoies: Flask, Django, FastAPI, REST API, GCP, AWS