Nishan Pantha

linkedin.com/in/nishparadox | github.com/NISH1001

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*

Experience

Graduate Research Assistant Jan 2022 – Present
UAH, NASA-IMPACT

- My primary focus is on Machine Learning research for earth science at NASA IMPACT
- Worked on large-scale data transfer solutions from ESA to NASA and developing open-source transfer solutions
- As a part of NASA's Bio-physical Science (BPS) collaboration, I worked on machine learning-based gene ranking algorithms for gene expression data from space-flown rodents and presented my poster for AGU 22. (Pending collaborative paper). I am currently extending this work for my MS thesis where I am working on NN-based feature attribution algorithms. Final Slide Deck, Poster Link

Senior Data/Research Scientist May 2021 – Dec 2021
Docsumo *Kathmandu, Nepal*

- Led, supervised, and implemented Machine Learning research and ML engineering for Document AI. This allowed me to do in-depth research on applied ML and work on the engineering side of building end-to-end scalable ML systems.
- Developed end-to-end training and inference pipelines within the Docsumo ecosystem. (This enabled anyone to use ML models to predict key-value pairs and tables)
- Developed a robust table extraction pipeline that drastically increased the "automatic extraction" of tables from documents. (This helped Docsumo to add TableML to the Docsumo ecosystem)
- Built several (extensible) internal tools to efficiently transfer/evaluate data within the Docsumo ecosystem.
- Built a table extraction engine/framework for extracting tables as per rules or ML models. This is in fact used in several production clients such as NDR.

Data Scientist Nov 2019 – May 2021
Docsumo *Kathmandu, Nepal*

- Implemented different NLP + Computer-Vision-based algorithms to assess any document's quality
- Built document fraud detection framework using images and texts
- Developed layout-based + graph-based models for key-value extraction

Co-Founder, Directory of Technology May 2018 – April 2019
MPercept Technology *Kathmandu, Nepal*

- Technically led and supervised all the projects related to ML.
- Helped various clients with building data-assisted systems.
- During this time, I also 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.
- Worked as a big

Big Data Consultant Feb 2019 – March 2019
Umniah Telecommunication *Amman, Jordan*

- As a part of MPercept Technology (Nepal), I traveled to Jordan and worked for Umniah Telecommunication for setting up a big-data system.
- Performed real-time data transfer from RDBMS to HDFS.
- Migrated CDR files to Hadoop system.
- Built backend for data analytics pipeline.

Part-time lecturer for AI May 2019 – July 2019
Janakpur Engineering College *Kathmandu, Nepal*

- Taught Artificial Intelligence course for final year B.E. Computer Engineering students.
- Slides: https://nish1001.github.io/engineering-ai/

AI Course Instructor Aug 2018 – Feb 2019

MPercept Academy *Kathmandu, Nepal*

- Taught, supervised, and mentored on Data Science and Machine Learning.
- Developed course modules for teaching ML topics like SVM, Decision Trees, Clustering algorithms, PCA, SVD, etc.
- Deeloped course modules for teaching Deep Learning concepts including CNN and RNN.

Software Engineer Nov 2016 – Feb 2018

Fusemachines *Kathmandu, Nepal*

- Worked on research projects related to chatbot and Intelligent Character Recognition (ICR)
- Built recommendation engine to propose use-cases for ML systems.

Data Science Intern Jun 2016 – Aug 2016

Phunka Technologies *Kathmandu, Nepal*

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

Projects

data-transfer-eval: tool for benchmarking data transfers | *Python* Jan 2022 – July 2022

- As a part of GRA under NASA-IMPACT, I helped develop a full framework to evaluate data transfer solutions which is to be made open-source soon.
- Repo to be made available: https://github.com/NASA-IMPACT/data-transfer-evaluation

BPS-numerical | *Python, XGboost, scikit-learn* July 2022 – Dec 2022

- As a part of GRA under NASA-IMPACT's machine learning research, I worked on developing initial benchmarks for space-flown rodent Liver RNA sequencing data.
- Developed a gene-ranking framework, **GeneRanker**, that uses ML models to identify topmost important genes contributing to certain physical attributes.
- Final Slide Deck, Poster Link

Docsumo: Document AI | *Python, Flask, Docker, Pytorch, Tensorflow, mlflow, spacy* Nov 2019 – Dec 2021

- Built most of the initial infrastructure for end-to-end Machine Learning Engineering. Eg: Table Extraction Framework, Graph-based key-value extraction frameworks, etc.

Damage Segmentation Detection | *Python, Pytorch, Tensorflow, scikit-learn*

- 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 damage segments.

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

- Worked on a temporal pattern-matching algorithm like Dynamic Time Warping
- Published at ICTAES
- Repo: https://github.com/NISH1001/humT

anuwadak: Statistical Nepali-English text translation, BE third-year project | *Python, numpy, etc.*

- Worked on Markov and N-gram models to predict texts.
- Repo: https://github.com/NISH1001/anuwadak

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

- My most popular open-source project that helps you find songs by name.
- Worked on crawler for songs+lyrics, caching mechanism, string matching algorithm.
- Developed item2vec model for song recommendations.
- Repo: https://github.com/NISH1001/playx

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

- My open-source tool to implement (from scratch) various mathematical animations such as L-Systems, simulations, etc.
- Repo: https://github.com/NISH1001/panim

tag-generator: open-source tool to generate tags from texts | *Python, numpy, matplotlib*

- Implemented TF-IDF-based algorithm to generate relevant tags for any texts
- Repo: https://github.com/NISH1001/tag-generator

Technical Skills

Languages+Technologies: Python, C/C++, Java, SQL, postgres, mongoDB, Docker, kubernetes, Google Cloud, AWS, S3
Web Frameworks: Flask, Django, FastAPI,
ML+NLP+CV Frameworks: scikit-learn, PyTorch, pytorch-lightning, Tensorflow, scikit-optimize, mlflow, open cv, mmcv, spacy, transformers
Data Science Tools: pandas, numpy, matplotlib, seaborn, plotly, Jupyter Notebooks