

Nishan Pantha

[linkedin.com/in/nishparadox](https://www.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 developed open-source transfer solutions
- Worked on applying machine learning algorithms on gene expression data for NASA BPS-numerical benchmark. (Pending poster for AGU 22 and collaborative paper)

Senior Data/Research Scientist

May 2021 – Dec 2021

Docsumo

Kathmandu, Nepal

- Led, supervised, and implemented Machine Learning research and engineering for Document AI. I helped build the core team for ML infrastructure ground-up. This allowed me to do in-depth research and work on the engineering side of building end-to-end scalable ML systems.
- Developed end-to-end generalized training, inference, and monitoring pipelines within the Docsumo ecosystem which are extensively used in production. This enabled users to use ML models to predict key-values and tables by allowing them to train on their own data.
- Built several (extensible) internal tools to efficiently transfer/evaluate data within the Docsumo ecosystem.
- Built a table extraction engine/framework for automatic extraction of tables using a robust combination of rules+ML algorithms. This helped Docsumo to add **TableML** to the Docsumo ecosystem. Users could annotate, train and predict tables. This is 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 from different documents like receipts, invoices, bank statements, etc.

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.

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...
- Developed course modules for teaching Deep Learning concepts including CNN and RNN.

Co-Founder, Mentor, Course Creator, Community Volunteer

Dec 2017 – Jan 2019

AIDevNepal

Nepal

- Organized 14 Saturdays workshop; taught and mentored more than 500 students in Data Science and ML.
- Developed course modules for different workshops.

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 data transfer benchmarks | *Python, rclone, Apache Nifi*

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 leverages ML algorithms to identify the topmost important genes contributing to certain physical attributes.
- Poster presentation for AGU 2022.
- Final slide deck:
<https://docs.google.com/presentation/d/1cEjAjpJ0ThlrJBfD1c83iotnwveuyqVoTteYrMOcaM/edit?usp=sharing>

Docsumo: Document AI | *Python, Flask, Docker, Pytorch, Tensorflow, spacy, mlflow*

Nov 2019 – Dec 2021

- Built most of the initial infrastructure for end-to-end Machine Learning.
- Eg: Table Extraction Framework, Graph-based key-value extraction frameworks, Document Fraud Detection framework, 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, final year project for BE | *Python, numpy, etc.*

- Worked on temporal pattern-matching algorithm such as Dynamic Time Warping.
- Published at: <https://ictaes.org/wp-content/uploads/2019/IJASS.01.02/2.11-16-Effects-of-Auto-Tuning-and-Pitch-Normalization-on-Query-by-Humming.pdf>
- Repo: <https://github.com/NISH1001/humT>

anuwadak: Statistical Nepali-English text translation, third-year project for BE | *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*

- 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

Data Science Tools: pandas, numpy, matplotlib, seaborn, plotly, Jupyter Notebooks