# Debojit Nath

debojit10.github.io | nathdebojit2000@gmail.com | linkedin.com/in/debojit-nath/ kaggle.com/denojitnath

## EDUCATION

**Indian Institute of Technology Madras**  2021 - 2023
*Diploma in Data Science*

**Techno India Institute of Technology**  2018 - 2021
*Bachelor of Computer Application, SGPA: 8.5*  *Kolkata, West Bengal*

**Kendriya Vidyalaya Fort William**  2009 - 2018
*CBSE, GPA: (class 12th) 86.5%, (class 10th) 95%*  *Kolkata, West Bengal*

## EXPERIENCE

**Data Science Intern**  November 2021 - February 2021
*Shortlyst AI*

- Conducted an assessment of the user's behavioral patterns to discern deficiencies within the system
- Developed LinkedIn data scraping bots and stored data in MongoDB database
- Optimized queries and created indexes to achieve 2M profiles/day speed
- Automated stats generation of bot performance for easy monitoring via Slack
- Fine-tuned DistilBERT model to predict industry tags for companies lacking available industry tags

**Data Science Intern**  June 2021 - December 2021
*Empliance Information Services India LLP*

- Developed scripts for data extraction process from more than 50 sources to create Structured Data for the legal team to use for Compliance Check using Selenium and BeautifulSoup
- Automated the process of downloading files of clients from MCA.gov, from logging in to saving them locally in a desired structure while handling captcha checking
- Fine Tuned HuggingFace model to classify Indian articles for Adverse Media Check with an accuracy of 95% and deploying an interactable webapp using Flask

**Data Science Intern**  February 2021 - May 2021
*Iha Consulting Services Pvt. Ltd*

- Built a Masked Face Recognition model using TensorFlow Object detection API and then converting to TFLite for optimized performance which performed better than out of the box solutions given they failed to detect faces

## RELEVANT COURSEWORK

**Deep Learning Specialization**  November 2019
*Certificate Link*  *Deeplearning.ai*

**Machine Learning**  August 2019
*Certificate Link*  *Stanford Online*

**Mathematics for Machine Learning: Linear Algebra**  February 2020
*Certificate Link*  *Imperial College London*

## PROJECTS

**Toxic Comment Classification Challenge**

- Participated in the Toxic comment classification challenge on Kaggle and trained a multi-label classification model which scored 0.9845 on public leaderboard

**Fashion Apparel classification**

- A dataset with imbalanced classes was used to train a VIsion Transformer model with re-sampling method to tackle the class imbalance problem and improved the accuracy of the Vision Transformer model from 66% to 69%

## TECHNICAL SKILLS

**Languages**: Python, C++, C
**Frameworks**: TensorFlow, PyTorch, Fastai, Flask
**Libraries**: Scikit-learn, HuggingFace, Keras, OpenCV, NLTK, Detecto, Selenium, BeautifulSoup