Indraneil Paul

Applied Scientist @ Amazon | CS Dual-Degree @ IIIT Hyderabad

androneil54@gmail.com
github.com/ineil77
ineil77.github.io
linkedin.com/in/neilpaul77
+91 8879073176

Work Experience

April 2020 – Present — Applied Scientist , Amazon, Bangalore

- Part of the Content Moderation Science team at Amazon Performance Advertising
- Experimented with multi-task multi-lingual pre-training regimes for Transformer-based networks
- Deployed an automated moderation solution as part of Advertising expansion to European marketplaces which increased automation by 28% while intercepting forbidden ads at a 17% higher rate
- Currently exploring multi-modal pre-training approaches to alleviate the operational burden of maintaining separate image moderation models

Python | Transformers | FastAI | PyTorch | CUDA C++

August 2019 – March 2020 — Software Development Engineer , Amazon, Hyderabad

- Part of the Transportation Plan Coordinator team for container shipments at Amazon Logistics
- Designed and partially implemented a constraint-aware cost-optimal package drop-off planner for allowing merchants to rank options to schedule last-mile drop-offs
- Scaled an tier-1 asynchronous service for year-end peak including database tuning, JVM latency optimizations, profiler based host type changes and scaling queue workers

Java | Spring | AWS SNS | AWS SQS | AWS DynamoDB | METIS

June 2016 – August 2019 — Research Assistant , Language Technologies Research Center, IIIT Hyderabad

- Worked under the guidance of professor **Ponnurangam Kumaraguru**
- Characterized the connectivity and activity levels of verified users on social media platforms
- Employed temporal, network and content based features for a discriminative model
- Uncovered the aspects of a user's profile, topic usage and activity that best predict verification
- Released a fully featured **dataset** of 400k+ users, containing 79+ million edges and 494+ million Tweets

Graph-Tool | FastAI | Neo4j | NLTK | Twitter API | Python | PoweRLaw | R

July 2018 – July 2019 — Research Assistant , Machine Learning Lab, IIIT Hyderabad

- Worked under the guidance of professor **Sujit Gujar**
- Explored the use of two-sided matching algorithms on dynamic graphs
- Researched location constraint aware graph matching
- Formulated applications of constrained mechanism design in resource exchanges and ride sharing

ParamILS | MATLAB | CVXOpt | Python | C++

May 2017 – August 2017 — Google Summer of Code , Green Navigation, Netherlands

- Improved fuel consumption prediction in electric vehicles given intended route and associated terrain
- Added bayesian optimization for optimal model hyperparameter selection for legacy models
- Modified pre-existing machine learning solution to leverage LSTM based time series prediction and reduced absolute forecasting error by 39%

TensorFlow | Pandas | BayesOpt | Python

Publications

Insights into, Analyses and Prediction of Verified Users on Twitter — Masters Thesis, IIIT Hyderabad
Indraneil Paul

What sets Verified Users apart? Insights, Analysis and Prediction of Verified Users on Twitter — WebSci 2019, Boston
Indraneil Paul, Abhinav Khattar, Shaan Chopra, Ponnurangam Kumaraguru, Manish Gupta

Elites Tweet? Characterizing the Twitter Verified User Network — LSGDA Worskshop, ICDE 2019, Macau
Indraneil Paul, Abhinav Khattar, Shaan Chopra, Ponnurangam Kumaraguru, Manish Gupta

Research Interests

- Natural Language Processing
- Self-Supervised Embedding Learning
- Self-Supervised Embedding Learning
- Meta Learning
- Multi-Modal Representation Learning
- Multi-Task Learning
- Domain Adaptation
- Knowledge Distillation
- Graph Neural Networks
- Optimization Methods
- Counterfactual Estimation

Education

2019	**Masters by Research**, IIIT Hyderabad	CGPA 8.52
2017	**Bachelors of Technology in Computer Science**, IIIT Hyderabad	CGPA 6.85
2013	**Higher Secondary**, PACE Junior Science College	HSC 89.1%
2011	**Secondary**, Hiranandani Foundation School	ICSE 90.2%

Capstone Projects

Wikipedia Search Engine
August 2017 - November 2017

Developed a search engine using term-document techniques capable of ranked retrieval over body and header text, references, links and metadata on a Wikipedia XML dump. Enhanced the system to support caching and B+ Tree indexing to enable the system to scale to the entire Wikipedia 14 GB corpus.

Python | BTrees | SAXParser

News Article Autosummarization
January 2017 - May 2017

Implemented an hierarchical LSTM based sequence to sequence model to automatically generate a grammatically coherent gist of a news article. Trained multiple sequence models in a topic aware manner aided by Hierarchical Dirichlet Process based multilevel topic models for sequence model selection.

Keras | Python

Kaggle Two-Sigma Financial Modelling Challenge
January 2017 - March 2017

Implemented a pipeline to compare the efficacy of various time series forecasting approaches in accurately predicting the future value of various anonymized financial instruments. Utilized GARCH models and multiple boosting techniques, eventually settling on an ensemble.

TensorFlow | Statsmodels | XGBoost | ARCH | Python

Author Context
January 2016 - May 2016

Created a system that ingests a large number of Computer Science research papers from the DBLP archives and uses the set of topical tags, abstract and title corresponding to each paper, to predict a field in which a certain author is likely to contribute in the near-future. Extended the system to predict future academic collaborations based on topical fingerprints. Used the memory-mapped key-value store LMDB to facilitate sub-second retrieval.

TensforFlow | LibSVM | LMDB | Python

NBA Result Prediction
July 2015 - December 2015

Developed a model that could predict, with competitive accuracy, the result of a basketball match between any two NBA teams factoring in player form, team form and past head-to-head results. Extended the model to factor in player synergies and team chemistry computed using a skills plus-minus framework.

MLPack | C++

Single-View Glare Removal
May 2015 - November 2015

Implemented a novel color-plane based approach to detect primary glare regions in images and rectify them using illumination constrained inpainting techniques extrapolating chromaticity and luminance from surrounding non-glare regions without using multiple images.

MATLAB | C++

Relevant Coursework

Graduate Advanced Computer Networks, Game Theory and Mechanism Design, Machine Learning, NLP Applications, Information Retrieval and Extraction, Statistical Methods in Artificial Intelligence, Modelling and Simulations, Computer Vision and Optimisation Methods

Undergraduate Data Structures, Algorithms, Operating Systems, Computer Systems Organisation, Distributed Systems, Database Management Systems, Computer Networks, Graphics, Digital Image Processing, Digital Signal Analysis and Applications, Number Theory and Cryptography, Natural Language Processing and Artificial Intelligence

Miscellaneous

2017	**College Football Team**, IIIT Hyderabad
2016	**Student Placement Coordinator**, IIIT Hyderabad