Akash Alok Mahajan

+1-(650)-546-5305 akashmjn@stanford.edu akashmjn.github.io LinkedIn/akashmjn

SUMMARY

Applied statistics: 3 years incl. work ex., courses/projects, TAs. Curious & enjoy wearing different hats.

- Teaching Assistant (TA) at Stanford for Machine Learning (CS229) and Deep Learning (CS230)
- Built a custom deep learning model on radio signals, being evaluated for deployment at SETI (Ongoing)
- Built an ECG annotation model comparable to inter-expert deviation on a public dataset 2016
- Initial data science team member at Tiger Global-funded smart vehicle startup in Bangalore 2015-16

Interested in learning to deploy ML/data products with related full-time roles. Graduating in June 2018.

Languages : Python, R, Scala, C, SQL, MATLAB
Software : Keras, Torch7, Apache Hadoop, Spark, AWS EMR/S3, Shiny, Processing

EDUCATION

Stanford University, Management Science & Engineering Stanford, CA
MS, Data Science track **GPA : 3.75** Sep 2016-June 2018
Courses : Small Data (MS&E226), Machine Learning, AI (CS229, 221)
Databases, Data Mining, Algorithms (CS145, 246, 161*), Intro. to Computer Systems (CS107)
Teaching Assistant : CS229 Machine Learning, CS230 Deep Learning* **(co-taught by Andrew Ng)**
(*Winter 2018)

Indian Institute of Technology, Madras Chennai, India
B.Tech., Chemical & Control Systems Engineering **GPA : 8.78/10** July 2011-July 2015
Courses/Projects : Modern Control Theory, Time Series Analysis, Kalman Filters

PROJECTS

Classification of ATA Radio Signals with SETI Institute+IBM Watson (CS341 Project) Stanford, CA
Identifying signal types from poor SNR data, Advisor : Prof. Jeffrey Ullman [report] Mar-June 2017

- Built an ensemble model on spectrograms : custom 12-layer CNN architecture, dynamic-programming based feature extractor. (Python/Keras)
- Model under evaluation to be deployed at SETI (6-class Accuracy 80%, 2-class F1 96%)

AI-based Music Generation (CS221 Project) Stanford, CA
Building generative models from MIDI data for conditional generation of melodies Ongoing

- Exploring the use of Markov chains, RNN based sequence models, and potentially Variational Autoencoders such as from Google Magenta for melody generation from MIDI files (Python/Tensorflow)

Building a Dynamic Memory Allocator (CS107 Project) Stanford, CA
Implementing `malloc`, `realloc` and `free` from scratch in C Mar-June 2017

- Implemented a segregated explicit free list, exceeding benchmark utilization and throughput targets

EXPERIENCE

Ather Energy, Data Scientist Bangalore, India
Building intelligence on smart electric scooters, part of the initial team of 2 Jul 2015-Jun 2016

Worked on initial feature roadmap with different internal teams, was presented at the product unveiling. Led 3 feature prototypes, contributed to infrastructure, and visualisation projects including :

- Systems to detect drivetrain damage, locate speed bumps, and profile riding styles from sensor data
- Infrastructure - CAN data parsers, initial Postgres schema, internal R/Shiny libraries
- Riding style visualization projects used to engage the early-adopter community [link]

INTERNSHIPS

Salesforce - Coolan (acquired in 2016), Data Science Intern San Francisco, CA
Datacenter hardware monitoring : Assisting Hadoop data-pipeline migration Jun-Sep 2017

- Built a data cataloguing tool for time-series backups in Amazon S3, and setup a pilot Spark+S3 cluster on Amazon Elastic Mapreduce (EMR). Learnt Scala, Spark and Hadoop tools over the summer.
- Built a pilot Spark ETL job to structure compressed JSON backups on S3, currently in use for migration.

Predible Health, Deep Learning Engineer Intern Bangalore, India
PoC for automated QT interval annotation of heart ECG waveforms using CNNs Jun-July 2016

- Built a custom 1-D convolution based CNN architecture on MIT-Physionet dataset
- Performance comparable to human inter-expert deviation on dataset (Mean +/- SD : 18 +/- 19.6ms)