NILESH PATIL

LEAD DATA SCIENTIST +917499680985

DETAILS

+917499680985
nilesh5760@gmail.com

SKILLS

- Machine Learning
- ML Systems
- Experimentation
- Deep Learning
- Data Visualization
- Apache Spark
- Dask
- Python
- R
- SQL
- Julia
- Scala
- Airflow
- Streaming Data
- Apache Kafka
- Cassandra

PROFILE

Experienced machine learning practitioner having extensive experience working with large datasets & complex systems. Exploring opportunities with scope to dive deeper into production machine learning systems at scale and pushing the envelope on applied ML in general. Playing a dual role of Lead Data Scientist & Data Science Manager.

Passionate about applied-ml at scale & building ML systems. Particularly interested in Computer Vision applications, Recommendation Engines & Time-series Forecasting systems.

Github | Medium | Linkedin | Website

EMPLOYMENT HISTORY

Lead Data Scientist at Dream11, Mumbai
April 2019 — Present

User churn detection at scale : Distributed time-series classification system

- Designed and led implementation of an end-to-end ML system from raw data to user-level churn detection on the scale of ~100M+ users using *Spark* and ensemble of models
- Automated A/B experimentation & performance tracking for different treatments in reducing user-churn

Real time forecasting system : Distributed time-series + sequence-to-sequence forecasting

- Conceptualized & led a team of data scientists to build a real time forecasting system for ~50k+ forecasts within 1 minute of system latency constraints (from raw data to actionable visual forecasts)
- Distributed inference system for real time processing & forecasting based on *Spark & Prophet*

Recommendation system : "Who to Follow" for social networks

- Led a cross-functional team to build a ML-system designed around *Spark + Airflow* with a 2-step recommendations process by formulating the problem as edge-prediction on large networks based on user interests & network properties

Image recognition & organization system for user generated content

- *Designed and implemented an end-to-end content detection & tagging system with Keras* based embedding generation & *FAISS* based retrieval at scale. System orchestration with *Airflow* & model deployments via *Docker* containers, *Sagemaker, Dask & Spark UDFs*

Featurestore : infrastructure for machine learning projects

- Conceptualized a distributed feature-store for ML systems and worked with multiple stakeholders across teams to bring it to production

Mentoring & Team Building :

- Mentored a team of Data scientists & Machine learning engineers to upskill for building machine-learning systems at scale
- Explored & refined skillets required to help build a high-impact Data Science team

Staff Data Scientist at Center for Vaccine Biology, University of Rochester, Rochester
December 2017 — March 2019

Bio-statistics & Computational Biology

- Led statistical computing initiative to bring statistical rigor in experiment design & computational analysis of hyper-spectral imaging data from experiments
- Implemented Python based automated analysis pipelines & ml-system to be used at the bio-imaging research labs at University of Rochester

Surface construction in live tissue from hyper-spectral microscopic scans

- 3D Surface and volumetric reconstruction from multiple 2D cross sections of live tissue using *Voronoi* mesh representations & 3D CNNs

Extracting uncertainty information from Deep Neural Networks

- Uncertainty information extraction & confidence interval building for deep neural networks using MC-dropout method & developed visualization apps for interpretation

Data Scientst at AXA, Pune
December 2014 — June 2016

- Predicting mortality rate across US counties using *AXA*'s historical claims datasets
- Scaling up statistical analysis pipelines using Spark & Python numerical computing ecosystem

Analyst at AbsolutData, Gurgaon
June 2013 — December 2014

Equipment failure prediction using sensor data

- Developed multi stage semi-supervised machine learning models for predicting equipment failure and triggering maintenance windows using sensor data & oil tests

EDUCATION

M.S. in Data Science, University of Rochester, Rochester
2016 — 2017

Analyzing large transportation networks

- Built large, dense, time-variant, multi-layer geo-spatial transportation graphs using NYC's transportation data & used Convolutional *LSTMs* to predict demand as node & edge properties of the network

Galaxy morphology classification using DNNs

- Collected & processed data from *Sloan Digital Sky Survey* (multi TB image dataset) using a mix of libraries for ETL, experimented with deep neural networks + skip-connections to predict hierarchical galaxy morphology classes

B. Tech., Indian Institute of Technology, Roorkee, Roorkee
2009 — 2013

PUBLICATIONS

- Prizant, Hen, Nilesh Patil, Seble Negatu, Alexandra Livingston, Scott Leddon, Andrew D. Luster, and Deborah J. Fowell. "***CXCL10+ perivascular clusters nucleate Th1 cell tissue entry and activation in the inflamed skin.***" (2020): 220-9 https://www.jimmunol.org/content/204/1_Supplement/220.9
- Patil, Nilesh, and Ajay Anand. "***Automated Ultrasound Doppler Angle Estimation Using Deep Learning***." In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 28-31. IEEE, 2019 https://pubmed.ncbi.nlm.nih.gov/31945837