

Nilesh Patil

2017903854 | nilesh.patil@rochester.edu | nilesh-patil.github.io | nilesh-patil | Nilesh Patil

Education

- Jun'16 - Dec'17 **University of Rochester**, M.S. in Data Science
Concentration in Computational & Statistical methods
- Jul'09 - May'13 **Indian Institute of Technology Roorkee (IIT)**, B.Tech. in Metallurgical & Materials Engg.

Projects

- Jun'16 - May'17 **Analyzing large transportation networks:**
- Built a large time variant network using NYC's transportation data (1 billion+ taxi trips), to analyze travel habits of residents and determining pressure points in the network
 - Used RNNs to predict demand at each node on dense, time variant geospatial transportation graphs with 2% rmse (*networkx, pandas, numpy, matplotlib*)
- Air quality prediction:**
- Collected & processed data collected by federal agencies across various open data portals
 - Trained Random Forest based ML regression model for predicting PM2.5 level in any given locality
 - Built multiple possible candidate models to contrast interpretability vs accuracy and achieved rmse of 6% (linear model), 2.5% (GBM) & 0.98%(Random forest) (*scikit-learn, pandas, matplotlib*)
- May'17 - Dec'17 **Machine learning driven Galaxy Morphology prediction:**
- Collected & processed data from Sloan Digital Sky Survey (multi TB image dataset) using a mix of SQL & python libraries
 - Built deep neural networks to infer detailed galaxy morphology for 10 million+ galaxy images & created a content based image similarity implementation (*tensorflow, keras, pytorch*)
- Extracting uncertainty information from Deep Neural Networks:**
- Implemented uncertainty information extraction & confidence interval building for deep neural networks with minimal network change, based on current cutting-edge research
- Time series analysis driven Exoplanet Detection:**
- Collected & processed time series data from Kepler Archive
 - Achieved classification accuracy of 70% with SVM, 85% using MLP & 92% for RNNs for Kepler's confirmed exoplanet database
- Extracting text from degraded ancient manuscripts:**
- Built a deep neural architecture for text extraction from heavily degraded ancient manuscripts

Professional Experience

- Dec'14 – Jun'16 **AXA**, Data Scientist/Sr. Analyst, Pune
- Worked with business partners to develop machine learning based predictive analytics frameworks
 - At AXA, we used Hadoop stack for storage + manipulation & R/Python/SPARK for analysis. Aggregation & Exploratory analysis was done using HIVE/IMPALA & R/Python
 - Underwriting pipeline & KPI visualization using tableau
 - Mortality rate error minimization using actual historical dataset from AXA's customers & transactions
 - Promoting Data science community by conducting training sessions, best practices meetups, machine learning & big data tools workshops for AXA-US & AXA-India
- Jun'13 – Dec'14 **AbsolutData Research & Analytics**, Analyst, Gurgaon
- **Sensor & transactional data analytics:** Developed multi stage semi-supervised machine learning driven model for prediction of engine failure in mining trucks
 - Developed random forest based predictive models for oil quality checks in heavy machinery
 - Implemented an unsupervised learning algorithm which helped reduce false alarms for the mining major based on actual effect of alarms based on historical alarms & maintenance data
 - Worked extensively hands-on as R & Analytics resource for the project using R, Hive & Hadoop. The data varied from Sensor data (13 TB) to Gigs of human entered - work order & maintenance data
 - **Marketing mix modelling:** Predictive regression model for optimal marketing expenditure

Skills

Programming

Python, R, SQL, Scala, Julia

Toolset

Numpy, Pandas, Scikit-learn, Tensorflow, Pytorch, Keras, plyr, dplyr, Pypark, graphX, HIVE/IMPALA queries, MapReduce using Python, Tableau, ggplot2, seaborn, matplotlib for visualization

Machine Learning

Experience in framing & solving business problems using machine learning techniques such as Random Forests, CNNs, RNNs, Graph convolutional networks, Support Vector Machines, GBMs, Linear Regression, Logistic Regression, Clustering Techniques (k-means, hierarchical clustering, knn etc), transfer learning