NITISH GOYAL

Houston, Texas | 215-605-2210 | nitish07g@gmail.com
https://github.com/Nitish76

EDUCATION

UNIVERSITY OF CALIFORNIA, San Diego **Jan-Dec 2018**
Masters in Algorithms and Data Structures

- Key courses: Graph Algorithms, NP-Complete Problems, String Processing and Pattern Matching Algorithms

UNIVERSITY OF HOUSTON, Texas **2016-Present**
Master of Science in High Performance Computing and Petroleum Engineering (3.6/4)

- Teaching Assistant for multiple courses; Awarded In-State Tuition on the basis of merit; GRE Score: 325/340

INDIAN INSTITUTE OF TECHNOLOGY DELHI, India **2009-14**
Master of Technology & Bachelor of Technology, Chemical Engineering (7/10)

- Awarded for Outstanding Performance in service to Student Community by Training & Placement Cell (2012)

DATA SCIENCE SKILLS AND EXPERIENCE

MODELING EARTHQUAKE DAMAGE IN NEPAL (For a Microsoft competition) **2018**
Placed in top quartile in competition for predicting building damage with a dataset of 20,000 observations

- Created a classification model to predict damage based on 38 features using Multi-class Decision Forest algorithm
- Optimized the model with hyper-parameter tuning to predict damage with a micro-averaged F1 score of 0.7
- Created a damage-assessment report including data visualizations (created with R ggplot) and recommendations

MICROSOFT CERTIFIED DATA SCIENCE PROFESSIONAL **2017**
Techniques employed include data pre-processing (integration, cleaning, transformation), exploratory data analysis, ML model improvement (feature selection, hyper-parameter tuning, cross-validation). Select projects listed below:

- Used adaboost tree algorithm to classify readmittance of diabetes patients (100,000+ points); AUC of 0.7
- Employed k-means clustering to find optimal clusters for adults based on census information (30,000+ points)
- Utilized random forest regression algorithm to predict bike demand (15,000+ points) with an R^2 of 0.94

ELIMINATING UNCERTAINTY IN MUD MOTOR BEHAVIOR USING DATA ANALYSIS **2016**
The project was aimed at reducing well drilling downtime by capturing relationship among mud motor parameters

- Data pre-processing: Cleaned (removed outliers) and normalized the data set (9000+ data points and 21 features)
- Exploratory analysis: Created a tornado chart of Pearson's correlation of variables to visualize interdependency
- Predictive modeling: Used artificial neural network with cross-validation to predict penetration rate

TOPIC MODELING OF NEWSPAPER ARTICLES USING NATURAL LANGUAGE PROCESSING **2018**

- Ingested and cleaned unstructured data (articles on Syrian war) from Global News database - Factiva using R(tm)
- Created document term matrix, term frequency histogram and word cloud to explore word patterns in the database
- Employed Latent Dirichlet Allocation using Gibbs sampling to classify articles into 5 groups using R(topicmodel)

TECHNOLOGY ENGINEER, RELIANCE INDUSTRIES, largest conglomerate in India **2014-16**

- Analyzed inspection data to create quality assurance reports for 30 columns (distillation, absorbers etc.)
- Identified troubleshooting areas based on analysis of various equipment and their performance metrics

SKILLS

Coding	Python (pandas, numpy, matplotlib, scipy, scikit-learn), R (ggplot2, dplyr), SQL, MATLAB, C++
Data Software	Azure ML, Tableau, Excel (power query, power pivot), RStudio (R Markdown), Jupyter notebooks
Mathematics	Probability and Statistics, Calculus, Optimization, Linear Algebra and Discrete Math
ML Algorithms	Random forest, Gradient boosting, Neural networks, Support vector machines, k-means clustering