KSHITIJ SHAH

kshitij.shah@rutgers.edu
kshitijshah.net
+ 1 848 239 8078
github.com/xitizzz

EDUCATION

Rutgers University, New Brunswick, NJ
MS in Computer Science, May 2018
GPA: **4.0**/ 4.0

Gujarat Technological University, India
BE in Computer Eng, July 2014
GPA: **8.58**/ 10

TECHNICAL SKILLS

Programming Languages
Extensive : Python, Java
Intermediate : C/C++, R
Basic : C#, Scala, Matlab

Data Science Skills
Extensive : SQL, Pandas
Intermediate : Spark, Plotly
Basic : Tableau, Hadoop

Machine Learning
Supervised Learning (Python, R), Deep Learning (Tensorflow/ Keras), Class Imbalanced Learning (imb-learn)

Other
Intermediate: Computer Vision (openCV), Web Design (HTML+CSS), OS (Linux/ Windows), Linux Shell
Basic: GPU Programming (CUDA)

COURSEWORK

Massive Data Mining and Deep Learning
Pattern Recognition (Unsupervised Learning)
Machine Learning
Introduction to Artificial Intelligence
Text Mining and Big Data Analytics
Data Structures and Algorithms
Database System Implementation
Algorithms I (Parallel Algorithms)
Computer Vision

ACCOMPLISHMENTS

- Won first prize in the Google games at Rutgers in 2016.
- Won first prize in 'GTU Codejam' a statewide coding competition in India.
- Won a prize for developing an extraordinary service app for 'txtWeb Planet of the Apps' competition.

EXPERIENCE

Data Scientist
May 2017 - Present

Center for Advanced Infrastructure and Transportation - CAIT, Rutgers University

- Lead data scientist on a risk analysis project for a major freight railroad network.
- Responsible for designing data flow, feature engineering, data analysis and visualization.
- Designed predictive models based on gradient boosted trees and neural network for rail defect prediction.
- Programmed an extensive data integration module to combine data collected from multiple sources with Pandas.
- Developed a web portal for data query and interactive visualization with Plotly.

Math Faculty
July 2015-July 2016

Academy for Training, Overseas and Management

- Worked as an instructor for GRE and SAT.

REAL-WORLD PROJECTS

Traffic Analysis from CCTV Footage (for CAIT)

- Developed a computer vision program to analyze the traffic from CCTV footage for Center for Advanced Infrastructure and Transportation.
- Computed count, relative speed, direction and time of crossing at a railway crossing for both pedestrians and vehicles.
- Achieved accuracy of 94% and computation time 60% less than actual video time, making it suitable for real time application.
- Wrote a modular program using object-oriented Python and openCV.

Electronic Health Record System

- Built a solution for storing, accessing and printing complex patient information for a medium scale hospital in India.

SELECTED ACADEMIC PROJECTS

Patent Data Exploration and Predictive Analysis using Spark-Scala

- Explored around 150 GB of approved patents and patent applications from USPTO data base using Spark to find patterns and insights.
- Built a classifier using SVM to predict the probability of patent application to be approved. Achieved around 80% accuracy.

Analysis and Visualization of Connections between Historical Figures

- Found similarities and connection between historical figures based on their Wikipedia page using NLP techniques with Python.
- Made a knowledge graph to visualize the results and interesting insights.

Crowd Simulation with Generative Adversarial Networks

- Deployed two individual models for simulating crowd movement behavior in public places, such as subway stations, in a team of four.
- Created a novel representation which used probabilistic heat map generated with overlapping Gaussian kernels.

Knowledge based Fake News Detection

- Developed a model with an information retrieval module and a feed forward neural network to integrate the knowdlegebase for fake news detection.
- Performed to exploratory analysis with Spark and Python.

PUBLICATIONS

Routing Algorithms with Distributed Network Table for Small World Networks | International Journal of Computer and Electronics Engineering