

# 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, August 2018  
GPA: 3.9/ 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

Text Mining and Big Data Analytics

Data Structures and Algorithms

Introduction to Artificial Intelligence

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

**Software Engineer  
Machine Learning  
- Intern**  
July 2018 - Present

**Data Scientist  
- Part Time**  
May 2017 - July 2018

### Black Knight Inc. - Philadelphia

- Designed and programmed a deduplication algorithm for scanned documents which can be deployed as independent service with minimum infrastructure requirement.

### Center for Advanced Infrastructure and Transportation - CAIT, Rutgers University

- Lead 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.

## 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.

## SELECTED ACADEMIC PROJECTS

### Confused Words Detection using Deep Learning

- Created a novel dataset using probabilistic models to simulate the behavior of typo and autocorrect. Labeled the confused words as result of autocorrect.
- Developed sequence tagging deep models (LSTM, GRU and Bi-LSTM) for the task detecting those confused words.
- Achieved the AUC score of 0.99 on the testing data.

### 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 knowledgebase for fake news detection.
- Performed exploratory analysis with Spark and Python.

### Patent Data Exploration and Predictive Analysis using Spark-Scala

- Exploratory analysis on 150 GB of patents and patent applications from USPTO data base using Spark, followed by predictive analysis using SVM.

## PUBLICATIONS

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