

# Sheallika Singh

✉ ss5136@columbia.edu | ☎ +1-347-279-5398 | 🏠 sheallika.github.io | 📷 sheallika | 🌐 sheallika

## INTERESTS

Data Analytics, Statistics, Machine Learning

## EDUCATION

### COLUMBIA UNIVERSITY

#### MS IN DATA SCIENCE

Sep'16 - Dec'17 | New York, US

Cum. GPA: 4.0/4.0

### IIT KANPUR

#### BS IN MATHEMATICS AND

#### COMPUTING

#### MINOR IN INDUSTRIAL ENGINEERING

#### AND MANAGEMENT

Graduated June'16 | Kanpur, IN

Cum. GPA: 8.5/10.0

## SCHOLARSHIPS

• Anita Borg Student Scholarship, 2015

• INSPIRE Scholarship, 2012-16

(awarded by Government of India)

## TEACHING

• Course Assistant: **Applied Machine Learning, COMS 4995**

• Course Assistant: **Advanced Analytic Techniques, QMSS 4018**

## KAGGLE PARTICIPATION

Bosch Production Line Performance

Bike Sharing Demand

Click Prediction

## SKILLS

### PROGRAMMING

• Python • C • SQL • Hadoop • Pig • Hive  
• Hbase • Spark • R • MATLAB • SAS • NumPy  
• Pandas • Scikit-learn • OpenCV • CUDA  
• Theano • Tensorflow • Jupyter • git  
• LaTeX • IBM System G • Linux • Windows

## COURSEWORK

### GRADUATE

Big Data Analytics

Data Analysis and Visualization

Machine Learning for Data science

Neural Networks and Deep Learning

Databases

Algorithms

Statistical Inference

Optimization

Learning Theory

Applied Probability and Statistics

Time Series Analysis

Regression Analysis

Real Analysis

Linear Algebra

Discrete Mathematics

## EXPERIENCE

### NETFLIX | DEEP LEARNING, COMPUTER VISION

May'17 - August'17 | Los Gatos, CA, USA

#### Perceived Emotion Recognition

- Tracking change of perceived emotion across a video, using a specially designed Recurrent Neural Network
- Tracing trajectory of emotions over the video using change in facial features, face points movement, background visuals along the video.

#### Dense Captioning Events in Videos

- Using the Dense Action Proposals to generate temporal proposals for actions
- Using attention mechanism and LSTM to generate dense captioning of the temporal proposals

### AGOLO | RELATION EXTRACTION FROM RAW TEXT

Natural Language Processing | Skills Used: Python

Jan'17 - May'17 | NY, USA

- Extract patterns of sentences in which input relationship (like Org-Founder) occurs from newspaper articles text data
- Extract newer Relationships for organisations which are not well known using these patterns, to develop Automated Information Retrieval model.

### FUZZY LOGIX | ENSEMBLE MODELLING FOR IN-DATABASE ANALYTICS

Data Science, Software Engineering | Skills Used: Python, SQL, R

May'15 - July'15 | NC, USA

- Applied bagging and boosting techniques to classification and prediction algorithms
- Observed significant improvements in accuracy (by 5-15%) for ensemble comprising decision trees ( Random Forest) and neural networks, but not for logistic regression
- Contributed towards development of packages for in-database analytics platform

## DATA SCIENCE PROJECTS

### REDUCING MANUFACTURING FAILURES [REPORT] [CODE]

Big Data Analytics | Skills Used: Spark, Python, R, IBM System G

- Utilized Extreme Gradient Boosting for feature extraction from 7 GB of data (with 4200 features)
- Random Forest (MCC Score: 0.40) and Gradient Boosting (MCC Score: 0.41) emerged the best models, to predict manufacturing defect

### IMAGE TO TEXT RECOGNITION FOR DEVANAGARI SCRIPT

NLP, Computer Vision, Machine Learning | Skills Used: Python, Caffe, OpenCV

- Devised algorithms using CNN and HOG for (samyutakshars) conjunct characters identification and separation into individual characters
- Decreased Word Error Rate by 10-15% than current state of art techniques

### CAPITAL BIKESHARE DEMAND PREDICTION [REPORT]

Data Mining, Data Analytics | Skills Used: R, Python

- Combined historical usage patterns with weather data to forecast bike rental demand in the Bikeshare program in Washington, D. C.
- Formulated random forests & exponential gradient boosting ensemble to predict bike rental demand with 88% accuracy

## OTHER PROJECTS

- **Statistical Market Research** : Modeled customer relationships using Statistical Analysis System to predict revenue band of customer with 94% accuracy
- **Forex Rate Forecasting Using ARIMA Model** : Concluded that linear model (ARIMA) prediction cannot beat random walk model
- **Neural Algorithm for Artistic Style Transfer** : Used Deep Convolution Networks (AlexNet and VGG-19) to fuse artistic styles and content from images
- **Predicting Recurrence of Breast Cancer** : Predicted recurrent cases of breast cancer using binomial logistics regression, with 72% accuracy