110 River Dr, Apt 2707
Jersey City, NJ 07310

ANUJ KATIYAL

anuj.katiyal@columbia.edu
(917)-770-6959

EDUCATION

Columbia University — **New York, NY** — **Sep 2016 – Dec 2017**

- Master of Science, Data Science, CGPA: 4/4
- Relevant Coursework – Probability and Statistics, Machine Learning for Data Science, Deep Learning & Neural Networks, Applied Machine Learning, Big Data Systems, Algorithms in Data Science, Storytelling with Data
- Course Assistant for Machine Learning (edX Columbia University MicroMasters Program) – Summer 2017 and Fall 2017

IIIT Hyderabad — **Hyderabad, India** — **Jul 2007 – Jul 2013**

- Master of Science (by Research), Computer Science, July 2013, CGPA: 4/4
- B.Tech. (Hons.) in Computer Science, May 2011. Dean's list in 3 out of 8 semesters for academic excellence. CGPA: 3.67/4
- Published research work at major Remote Sensing Conferences and was awarded the **Best Young Scientist Presentation Award** for research presented at ISRSE35 conference held at Beijing, China in April 2013
- Teaching Assistant - IT Workshop, C Programming, Computer Vision and Introduction to Databases (Head TA)

SKILLS

Python (IPython, Numpy, PySpark, Scipy, Pandas, Matplotlib, NetworkX, Scikit-Learn, Keras); SQL; R; Spark; AWS; Presto; Hive; JavaScript (jQuery, D3.js), Web Development (Python - Flask), Mobile App Development (iOS - Swift)

DATA SCIENCE PROFESSIONAL EXPERIENCE

Condé Nast, NY — **Data Scientist** — **Feb 2018 – Present**

- **Falcon – Virality Prediction for Content Data** *(Tools Used: Python, PySpark, Presto, AWS, Airflow, Apache Kafka)*
  - Implemented an end to end AI tool which discovers trending content based on real time traffic, predicts how they perform on social media using uplift models and recommends top content to share on social media.
  - The tool has been rolled out on all 14 Condé US Brands for FB and Twitter social platforms. Some of the major brands include Vogue, New Yorker, GQ, Wired, Architectural Digest, Condé Nast Traveler and Bon Appetit.
  - Falcon drives around 30% of all FB traffic over 14 Condé US Brands, around 70% social traffic on evergreen content for US brands and has resulted in 30% YOY increase in FB interactions (likes, shares and comments) for Q4 2019.

Twitter, NY — **Data Science Intern** — **Jun 2017 – Sept 2017**

- **Media Data Acquisition Module for PyCX Library** *(Tools Used: Python, PySpark, Scala, Scalding)*
  - Improved PyCX (the PySpark framework for Data Analysis) with increased Twitter data source access. Further, added PyCX Media Sampler to acquire media data using Scalding Jobs (Scala) and PySpark notebooks for ML/DL tasks
  - Twitter Toxicity Intern Challenge – Supervised Classification challenge for the detection of toxic tweets. Used ML, NLP and Deep Learning methods and stood 2nd amongst the Science Org Interns.

TCS Innovation Labs, India — **Researcher – Applied Data** — **Aug 2013 – Aug 2016**

- **Analyzing GitHub Experts** *(Tools Used: Python, NumPy, Pandas, Scikit-Learn, NetworkX, D3.js, Flask, AWS)*
  - Analyzed the code data of several JAVA experts to discern their skillset, tools and techniques used most often and ways to tag and classify these into subfields for visualization.
  - Publication – Saxena R., Katiyal A. and Pedanekar N.; "I Know What You Coded Last Summer: Mining Candidate Expertise from GitHub Repositories", CIKM Workshop on Data-Driven Talent Acquisition, Oct 2016

ACADEMIC DATA SCIENCE PROJECTS

- **Estimating Taxi Demand at LaGuardia Airport** *(Tools Used: Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras)*
  - For the Capstone Project, worked with NYC's Taxi and Limousine Commission (TLC) to predict hourly estimates of taxi pickups at NYC LaGuardia (LGA) Airport, to address the demand-supply issue of taxis at the LGA airport.
  - Best regression model included a Tree-based ensemble achieving a Mean-Absolute Error (MAE) of 56.9 and R^2 of 0.908. Also used a Long Short-Term Memory (LSTM) model, which achieved an MAE of 48.1 and R^2 of 0.921.
- **Data Projects** *(Tools Used: Numpy, Pandas, Scikit-Learn, NetworkX, D3.js, Flask, AWS)*
  - Stood 2nd amongst 100 teams participating in In-Class Kaggle Classification Analysis of Bank's Marketing Campaign to analyze Subscription Status. Achieved best ROC-AUC score of 0.798 using ensemble methods.
  - Quantified Self: Analyzing Personal Garmin Data http://bit.ly/2vdiox7
  - Storytelling Project - Marvel vs DC Superheroes compared using D3.js. http://bit.ly/2wlvZAe

HOBBIES

- Quantified Self Enthusiast, Avid Runner, Blogging (https://anujkatiyal.com)