



AKSHAY AGARWAL

23 Pine Street | Morristown, NY 07960 | +1 607-262-4883 | aa2657@cornell.edu |  

SUMMARY

A distinct blend of Software Engineer and Data Scientist

A highly motivated and passionate individual with three years' experience as software engineer, eager to apply rigorous scientific methodologies on complex datasets. Experienced in ML Algorithms, Big Data Tools, AWS and Web Scraping

EDUCATION

CORNELL UNIVERSITY

Ithaca, NY

Master of Professional Studies in Applied Statistics (Data Science Concentration), GPA: 4.1

Aug 2018- May 2019

- **Selected Coursework:** Machine Learning • Big Data • Language Processing • Database System • Linear Models

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE (BITS), PILANI

Pilani, India

M.Sc. (Hons.) Mathematics & B.E.(Hons) Electronics & Instrumentation

Aug 2010-May 2015

- Research Intern at Intel Labs, Intel (India) in Fall 2014 and University of Toulouse (France) in Summer 2013

SKILLS & LANGUAGES

- **Programming Languages:** Python, JAVA, SQL, R, Perl
- **Tool/Technologies:** Apache Spark, Hadoop, HDFS, Unix, AWS EC2, AWS S3, Hive, Git
- **Python Libraries:** Scikit-Learn, Pandas, TensorFlow, Keras, NetworkX, Scrapy, Selenium, NumPy, NLTK, Matplotlib

EXPERIENCE

KMK Consulting *NJ-based pharma consulting firm*

Morristown, NJ

Data Analyst, R&D Team

July 2019- Present

- Worked on big data project for Top 3 pharma company, Developed KPI's to identify key opinion leaders in Oncology and rank 10k Oncologists by their influence on social media
- Gathered, merged and cleaned six data sets obtained from online sources including Twitter, YouTube and US News to construct a network analysis model; Recognized by hierarchy for excellent Python skills and offered full time role

FCB Healthcare *NYC-based healthcare advertising firm*

Ithaca, NY

Consultant, Opioid Drug Prescriptions

Jan 2019- May 2019

- Performed Statistical Analysis in Python as part of *capstone project at Cornell* to determine the association between physician's personality trait and clinical treatment preferences using data of 450K physicians in Python
- Implemented PCA to deduct features, and developed models using Linear Regression, Gradient Boosting and XGBoost

FAIR ISSAC CORPORATION (FICO)

Bengaluru, India

Software Engineer, Scores and Predictive Analytics (*Received early promotion*)

Jun 2015- Jul 2018

Interfaced with financial sector clients to develop and deploy end-to-end customized predictive credit scoring platforms using JAVA

- Worked closely with data scientists and global clients to define business problem and analyze software requirements; researched machine learning techniques and implemented best algorithms to developing scoring models
- Created Spark application in JAVA to preprocess 1 TB of daily telecom data under 20 minutes for Philippines largest telecom company; extracted data from AWS S3 and deployed the application on AWS EC2
- Proactively provided client support, quarterbacking team in resolving critical production issues and saving ~\$3M in contracts
- Mentored two interns on module development in Perl; Received Spot award in 2016 for exemplary performance

SELECTED ACADEMIC PROJECTS

CORNELL UNIVERSITY

Ithaca, NY

Wine Recommendation System ([Link](#))

Feb 2019 – May 2019

- Developed app to parse database of 132k wines and suggest recommendations based on user input (price, grape type, region)

Wikipedia Clustering Competition

Mar 2019- May 2019

- Grouped 11k Wikipedia articles in four clusters using raw text files and directed graph matrix to connect various articles through hyperlinks using *unsupervised learning algorithms* in Python
- Performed PCA, SVD, Spectral clustering and K-means to achieve accuracy of 98.5%; landed in top 5% of the class

Device Classification for Trump's Tweet (In-class Kaggle Competition)

Sep 2018 - Dec 2018

- Predicted whether tweet was sent by Trump or his office and performed sentiment analysis using 2000 pieces of text
- Achieved 85% accuracy by implementing K-cross validation and Random Forest in Python, landing in top 20% of the class

Big Data Text Analytics using Hadoop and MapReduce Framework

Sep 2018 - Dec 2018

- Wrote MapReduce functions in Python and used Hadoop Streaming for computing word frequency statistics of large text files

AWARDS

- Earned Merit Certificate for achieving 99.9 percentile in India-wide high school Mathematics exam in 2007
- Awarded Parikh Parivar Prize for All India Actuarial Common Entrance Test (1st rank amongst 3,000 candidates) in 2015