MS in Data Science Columbia University

TIAN ZHENG

ASSOCIATE DIRECTOR FOR EDUCATION, DATA SCIENCE INSTITUTE

PROFESSOR OF STATISTICS

COLUMBIA UNIVERSITY

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

DSI Education

Degree:

Master of Science   2014

Non-Degree:

Certification   2013

ColumbiaX:

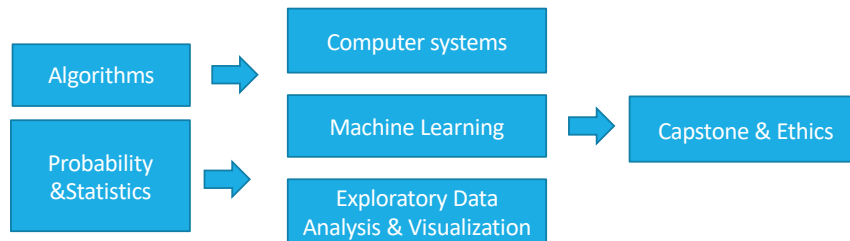Online Courses   2016

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

MS Curriculum - Required Courses

◦ Algorithms for Data Science
◦ Probability Theory + Statistical Inference & Modeling
◦ Exploratory Data Analysis & Visualization
◦ Machine Learning for Data Science
◦ Computer Systems for Data Science
◦ Data Science Capstone and Ethics

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Flexible curriculum

◦ Data Science students came with diverse background
  ▪ Engineering: CS, OR, EE, etc
  ▪ Stat/Math
  ▪ Other: Social sciences, Journalism, etc
◦ Required courses form a solid foundation in DS
  – seven required courses, three elective courses
◦ Alleviate the situation by allowing formal waivers and substitutes.

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Selected Electives

Machine Learning Products for Data Science

Applied Machine Learning

Translational Bioinformatics

Causal Inference for Data Science

Sustainability Technology and the Evolution of Smart Cities

Information Processing: Big Data Analytics

Applied Machine Learning for Financial Modeling and Forecasting

Deep Learning for Computer Vision, Speech, and Language

Big Data in Finance
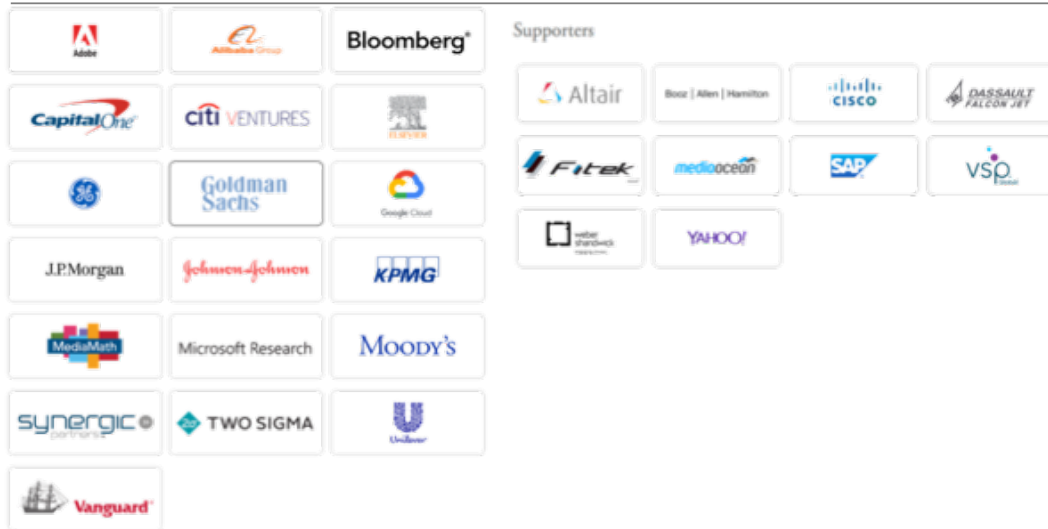
Applied Machine Learning for Image Analysis
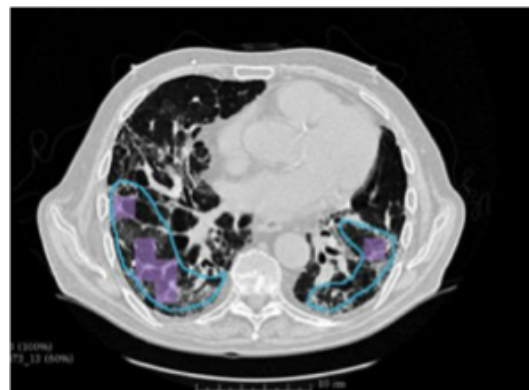
Computational Models of Social Meaning

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Capstone Projects

STUDENTS WORKING WITH INDUSTRY AFFILIATES

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

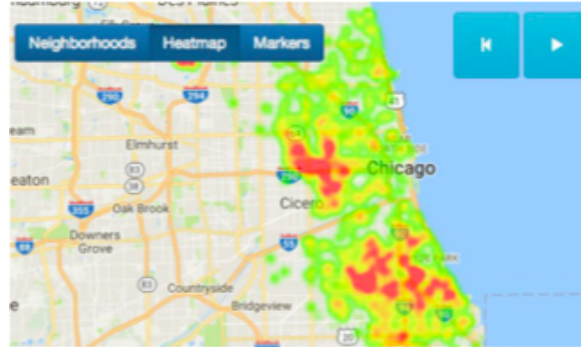# Industry Affiliates Program

Supporters

Teaching Computers to Spot Damaged Lung Tissue

o Picking out abnormal lung tissue in high-resolution CT images of the lung takes time and skill.

o This project was proposed by General Electric Global Research.

o The team built a model for predicting whether high-resolution images of lung tissue showed evidence of scarring, or pulmonary fibrosis.

Visualizing Gun Violence and Poverty in American Cities

- A project for Cure Violence and the consulting firm Booz Allen Hamilton.
- This team combined gun-crime and census data for New York, Chicago and Baltimore, to understand the interplay between crime, poverty and other factors.
- An interactive tool they built lets users explore the data and put several trends in sharp focus:
  - Chicago gun-crime peaks annually in July and tapers off in February;
  - New York had the lowest gun-crime rates of the three cities, possibly due to its well-funded gun-prevention program;
  - and gun crime in Baltimore has been steadily rising

The above heat-map shows how pockets of crime in Chicago shifted to different neighborhoods over a six-year period.

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Automating Case Law Analysis

- Lawyers must understand legal precedents for plotting trial strategy and predicting trial outcome.

- This team used case law from the U.N. Office on Drugs and Crimes.

- Developed an interactive system for analyzing data on legal precedents.

Main dashboard filtered for cases where there was criminal intent

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

DSI Feeds New York City's Start Up Community

Agolo algorithmically curates your twitter feed

Wordseye lets anyone create 3D images

eBrevia uses machine learning to summarize legal documents

Fero Labs brings machine learning to factories to optimize production

TextIQ uses natural language processing to change the consumption of big data

Droice Labs uses artificial intelligence to predict how drug treatments affect patients

Chipscan, a cybersecurity firm, identifies malice in chips

Vidrovr analyzes video collections to make them searchable

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Data Science Student Employers

AIG, Adhoc, Alvarez & Marsal, **Amazon**, American Express, Amgen, Amper Music, Amphora, Audible, Barclays, BCT Partners, **Blackrock**, Capital One, **Capital One Labs**, CartoDB, CKM Advisors, Clipper Data, Collibra, Comcast, Creative Chaos, **Crisis Text Line**, Deloitte, Digital Reasoning, Droice, Early Signal, eBay, EMC Corporation, eScience Institute at U of Washington, **Facebook**, Factset Research Systems, FarePortal, **FDNY**, GE Research, Glassdoor, **Goldman Sachs**, **Google**, Guy Carpenter, Handy, Hover, **IBM**, IBM Social Good Fellowship, Intersection, Intuit, Jet.com, Jobdiva, Kora Capital Management, Manhattan DA's Office, **McKinsey**, MediaMath, **Merck**, **Microsoft**, Milliman Max, MoneyLion, **Morgan Stanley**, Mount Sinai, MSCI, Mylan Pharma, **NBCUniversal**, Nestle Waters, **Netflix**, NYC Department of Buildings, OnDeck, **Palantir**, Paypal, Pfizer, Pfizer, Pixel Place, **Point72**, Primus, Quaera, Quarterspot, RBC, Red Ventures, Resolvity, SAP, Satmap, Singapore Bank, **Spotify**, Spreemo, Springbot, Swiss Re, Synergic Partners, TAPAD, The Hartford., Tomorrow Networks, Trans Org Analytics, Tremor Video, Trifecta, Trinnacle Capital, TuneIn, **Twitter**, **Uber**, Uncommon Schools, **United Nations**, Venus Tech Ventures, Verisk Analytics, Viacom, Vulcan, Walmart, Yelp.

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Top Data Scientist Jobs:
Data Scientist
Business Analyst
Data Analyst
Statistician
Senior Data Scientist
Chief Scientist
Research Scientist
Analytics Manager
Senior Business Analyst
Analytics Consultant
Business Intelligence Consultant
Data Architect
Research Analyst
Chief Data Scientist
Director of Analytics
Quantitative Analyst
Senior Web Analyst
Lead Analyst
Entrepreneur

100% Placement of Inaugural Class

Thanks!

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK