

JAMES BLAU

DATA SCIENTIST

✉ jamesblau@gmail.com
☎ (203) 246-1925
📍 Weston, CT 06883
🌐 linkedin.com/in/james-blau/
📷 jamesblau

Moved into data science after working in software development, making "big data" applications using Hadoop/Spark/Scala, and before that performing MRI research, making radio antenna arrays.

Skills

PROGRAMMING LANGUAGES

Scala
Python
SQL
Bash

DATABASES

MongoDB
HBase
PostgreSQL
Elasticsearch

MACHINE LEARNING

Linear/Logistic Regression
Tree-based methods (CART)
Neural Networks
Ensemble models
Dimensionality reduction
K means

NATURAL LANGUAGE PROCESSING

Topic Modeling (NMF/LDA)
Collaborative Filtering
TF-IDF
Conditional Random Fields

DATA TOOLS

Spark
HDFS
Pandas
Numpy
Sklearn
Keras

DATA VISUALIZATION

Matplotlib
Seaborn
Tableau

OTHER

AWS
Bash/CLI
3D Modeling/Printing
RF Engineering

Data Science Projects

Next Up: Movie Recommendation System

- Used topic modeling (spaCy, Gensim) to perform natural language processing of movie scripts and reviews
- Combined content-based recommendation using these topics with collaborative filtering (SVD) using online review scores to build a movie recommendation system

Game!: Image Recognition for Esports Statistics

- Trained a convolutional neural net to recognize character actions in an esports fighting game to generate game statistics from video
- Built a pipeline with Matplotlib and opencv2 to produce labeled data used to fine-tune VGG16 with Keras

Alzheimer's Diagnosis with MRI Data

- Used gradient boosting with decision trees to predict Alzheimer's diagnoses
- Combined data from MRI brain scans with cognitive impairment test (MMSE) results to improve recall compared to using MMSE alone

Biff! Bam! Pow!: MMA Outcome Prediction Model

- Built a LASSO regression model predicting outcomes of Mixed Martial bouts based on attributes and past fights of each contestant
- Web-scraped using BeautifulSoup to build and parse a large database of fighter and statistics

Optimizing Volunteer Deployment with MTA Ridership Data

- Analyzed MTA ridership data to determine optimal placement of volunteers to drive engagement and fundraising
- Performed extensive data cleaning, EDA, and data visualization with seaborn to make sense of an inconsistent dataset

Experience

Metis

Data Scientist

- Metis is an accredited 12 week immersive data science bootcamp focused on Python, statistics, supervised and unsupervised machine learning, data analysis, visualization techniques and database management
- Designed five end-to-end projects utilizing data acquisition, cleaning, and modeling to derive data-driven insights, presenting results to technical and nontechnical audiences

New York, New York
Mar. 2020 to Current

Applerouth Tutoring Services / Private Tutor

Math + Science Tutor

- Tutored middle school and high school students in math and physics, communicating high-level concepts in an easily understandable way
- Developed personalized programs to build executive-functioning and broadly-applicable academic skills

Weston, CT
2012 to 2013, 2018 to 2020

BookNewClients.com

Software Consultant

- Designed a marketing automation platform for wealth managers, offering customized financial education content to prospective clients
- Deployed personalized splash pages to drive engagement, capturing leads and delivering them to the participating wealth manager

New York, New York
2017 to 2018

Tresata LLC

Software Engineer

- Developed analytics programs for finance and retail, using Map-Reduce frameworks (such as Spark and Cascading) and other distributed big data tools in the Hadoop ecosystem (such as Kafka and Elasticsearch)
- Applications included entity resolution, fraud detection, and product recommendation systems
- Researched, tested, and implemented new algorithms and tools, including graph algorithms such as graph clustering and Conditional Random Fields, and probabilistic data structures such as Count-min Sketch and MinHash, to work with data at scale
- Filled a variety of roles in a startup environment, from administering Hadoop clusters, writing deployment software, and conducting cryptographic license-checking, to working directly with customers, core product development, and some front-end design

New York, New York
2013 to 2017

Martinos Center for Biomedical Imaging, Massachusetts General Hospital

MRI Researcher

- Used 3D printing, CAD and network analysis to design and construct transmit/receive coils and other hardware for MRI, including a 64-channel "Connectome Coil" used in the Human Connectome Project
- Second author on a paper accepted by the ISMRM's journal, Magnetic Resonance in Medicine, and co-author on accepted submissions to ISMRM's 2010, 2011, and 2012 conferences

Boston, MA
2009 to 2012

Publications

A 64-channel 3T array coil for accelerated brain MRI (second author)

Magnetic Resonance in Medicine, Vol. 70, No. 1, pp. 248-25, <https://www.ncbi.nlm.nih.gov/pubmed/22851312>

Discover Magazine: Meet the World's Most Advanced Brain Scanner

Education

Harvey Mudd College
Studied Physics

2006 to 2009