JOSHUA BANKS MAILMAN PH.D. DATA SCIENTIST

Joshuabanksmailman@gmail.com 917-945-1854 New York, NY 10032 JoshuaBMailman /in/joshua-banks-mailman/ jmailman

SKILLS

DATA SCIENCE: Python, Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Scipy, NItk/NLP, Beautiful Soup, Excel, Sql, Linear Regression, Cross validation, Regularization (Lasso & Ridge), Dimension reduction (PCA/SVD), Classification, Unsupervised machine learning, Hypothesis testing, Data cleaning / pre-processing / reshaping, Extreme Gradient Boosting (XGB), Web scraping, Topic modeling with NMF/LDA, TF-IDF, Feature engineering, Oversampling, Word2Vec, Streamlit, Data smoothing, NN image recognition, Java, Javascript, C++, SAS, iOS/Objective C

SOUND, SENSORS, INTERACTIVITY, AUDIO-VISUAL, NETWORKING TECHNOLOGY: Max/MSP, Leap motion controller, RTcmix, Open Sound Control (OSC) protocol, Screenflow, OSC-skeleton, OpenNI/NITE, Processing (java-based graphics programming), Audacity

EDUCATION

Eastman School of Music, University of Rochester
Ph. D. Music Theory 2010
Awarded Professional Development travel grant for Yale Summer School course on Arnold Schoenberg's music , Vienna, Austria (2001)
Honorable mention, Jerald Graue award for excellence in musicology (2002)

University of Chicago
A. B. Philosophy 1995

Baruch College, CUNY
Structured Programming in C 1995
Object Oriented Programming in C++ 1996

General Assembly, NYC
Part-time, certificate Javascript 2017

Harvestworks, NYC
Part-time, certificate Generative Graphics Programming with the Java Processing Framework 2011

Columbia University
Audited Advanced Topics in Computer Music (RTcmix/MaxMSP/iOS) 2011
Awarded grant to study Gagaku in Tokyo for 5 weeks
Awarded the Mellon, Woodrow Wilson fellowship for Humanistic Studies

EXPERIENCE

Metis, *Data Scientist*, New York, NY — Jan. 2021 - Mar. 2021
- Completed an immersive 12-week data science bootcamp with a strong emphasis on project-oriented skill-building in problem solving, data wrangling, statistical modeling, machine learning, and communication of deliverables
- Project highlights below

Columbia University, *Adjunct Assistant Professor of Music*, New York, NY — 2010 - 2019
- Devised, engineered, demonstrated, and performed-with several interrelated algorithm-driven interactive systems: 1) iOS/iPhone custom app that controls a music generating algorithm by tracking hand tilt through the iPhone's accelerometer, 2) Two multi-sensor full-body interactive systems that used the Kinect infrared video camera and custom-built wireless sensor gloves to steer over a dozen continuously streamed parameters into self-designed algorithms which generated music and live animated graphics simultaneously, which were projected in spatialized audio and to a viewing screen for an audience in realtime, 3) An interactive algorithmic music system controlled by foot pedals and the Leap Motion infrared hand-tracking sensor
- Wrote and achieved publication of 11 academic articles in top-ranked academic journals that are blind-peer-reviewed, in addition to more than 10 other articles, 5 book chapters, and numerous newspaper/magazine/blog items. These have been read, assigned, and taught at numerous universities and cited and quoted in numerous articles, dissertations, and books
- Gave over 35 distinct original-research peer-vetted lecture presentations and demonstrations on over 50 occasions at competitive conferences and symposiums across North America, South America, Europe, the United Kingdom, and Japan.
- Taught Music Theory and Music Humanities for 2010-12, 2013-15, and 2017-19
- Supervised teaching assistants (2010-12)
- Created numerous instructional videos and pedagogical technologies for use in and outside the classroom
- Earned a grant from the Columbia University Institute for Medieval Japanese Studies study Gagaku music (performance and theory) in an intensive summer-long program in Tokyo (2014)
- Appeared on national television on ABC News Nightline as an expert on emotional response to pop songs: http://tinyurl.com/lymasm6

University of Alabama School of Music,, *Full-time instructor of Music Theory*, Tuscaloosa, AL — 2015 - 2017
- Taught specialized seminars and review courses to graduate students and undergrads

University of California Santa Barbara, *Visiting Assistant Professor of Music Theory*, Santa Barbara, CA — 2012 - 2013
- Was invited (hired) to lecture on and demonstrate self-designed interactive audio-visual sensor-driven algorithmic performance system at Ground-breaking Research/Innovative Technology (GRIT) talks of the UCSB Research Mentorship Program / UCSB Summer Cultural & Enrichment Program
- Achieved publication of 'Improvising Synesthesia: Comprovisation of Generative Graphics and Music' in Leonardo Electronic Almanac vol. 19(2): Live Visuals
- Taught specialized seminars and review courses to graduate students and undergrads, and coordinated/supervised teaching assistants

New York University, *Adjunct Instructor of Music*, New York, NY — 2010 - 2012
- Awarded travel grant to present research at the Intl. Society for Music Perception and Cognition conference in Thessaloniki, Greece
- Taught basic harmony and ear-training to undergrads and music criticism/writing to graduate students
- Oversaw the grading process and supervised those grading the work

JOSHUA BANKS MAILMAN PH.D. DATA SCIENTIST

PROJECTS

Experiments in captioning, 'Wit a twist: The Amusemater Captioner — Mar. 2021 - Mar. 2021

- Created a web app that lets machine learning dip its toes into the waters of witty wordplay by inventing image captions that make you smile.
- Used a Neural Network image classifier joining hands with NLP to algorithmically dance right to the verbographic punchline.
- Data: 5500 images from IMAGENET, 1500 English idioms from https://7esl.com, and corpuses: SentiWordNet, Gensim Text8, and GoogleNews
- Pretrained Neural Network (Xception) used for image classification, Gensim Word2Vec word embedding used to identify words related to the image label, Scikit-learn to make TF-IDF correlation matrix between SentiWordNet corpus and IMAGENET labels to identify words related to the image label, NTLK, and other NLP packages (Pronouncing, Phonetics, and English-to-IPA) to identify similar sounding words (rhymes, assonances, etc.)
- Result: A Streamlit app that takes any image, classifies its content, and algorithmically invents a (witty) caption indirectly relating to the image's content

New ways to read 'Language on a Holiday' — Feb. 2021 - Feb. 2021

- Used unsupervised Natural Language Processing (NLP) machine learning to distinguish individualized prose style and varying topics in philosophical texts
- The texts were three complete books: Hume's Enquiry Concerning Human Understanding, James's Pluralist Universe, and Whitehead's Process and Reality drawn from Gutenberg.org
- Leveraged NLTK, Scikit-learn, SpaCy, Numpy, Pandas, Matplotlib, the algorithms deployed include: Non-negative Matrix Factorization (NMF) topic modeling, feature-engineered (modified) TF-IDF (term frequency - inverse document frequency) matrix to model prose style, Scatterword and spline graph and visualizations
- Result: A computational method to distinguish prose styles and to isolate self-explanatory and idiosyncratic (eccentric) prose passages

To cocktail or not to — Jan. 2021 - Feb. 2021

- Built a classification model to identify what food ingredients can be used in a cocktail
- Data was acquired from the CocktailDB.com JSON database via web-scraping with BeautifulSoup, and from the FoodDB.ca relational-database , which listing concentrations of 15,000 food compounds
- Data was preprocessed and merged using NTLK (Levenshtein distance), Pandas, and Dask
- Using Scikit-learn various classifiers (Random Forest, KNN, Naive Bayes) were attempted but ultimately Logistic Regression and XGBoost (extreme gradient boosting) were used
- Class imbalance (7:1) was addressed with both oversampling and by using alternative performance metrics log-loss cross-entropy and ROC-AUC
- 8000 additional food features were engineered
- Principal Component Analysis (PCA) was used for dimension reduction
- Visualization using Matplotlib and Seaborn
- Result: Potential ingredients for cocktails were discovered, such as thyme-like winter savory, cucumber-like tinda squash, and Mexican mamay fruit

Scrape flix — Jan. 2021 - Jan. 2021

- Built a regression model to predict IMDB ratings of films by 7 auteur directors (Hitchcock, Godard, Scorsese, etc) from metadata
- BeautifulSoup was used to scrape metadata from IMDB, resulting in over 9,000 datapoints
- Polynomial and logarithmic features were engineered
- Using Scikit-learn, linear regression with lasso and ridge regularization and cross-validation were applied
- Seven linear regression models were developed, one for each director. Mean Absolute Error and R^2 were were used as metrics.
- Result: Of the directors considered, Hitchcock's film ratings are predicted most reliably (R^2=.8), based on year of release, ratings count, and their logs