DR. KOBI ABAYOMI drkobiabayomi@re-search.website

DATA SCIENCE EXECUTIVE, AUTHOR, AND CONSULTANT IN HIGH DIMENSION DATA, MACHINE LEARNING & ARTIFICIAL INTELLIGENCE, MEASURES OF DEPENDENCY & DETERMINISTIC/STOCHASTIC OPTIMIZATION. I have more *than 20 years of experience leading Data Science teams to consult, publish & produce.* I created and led Data Science at Warner Music Group where my team and I instantiated models which accelerated music demand & asset monetization. I created and led the Data Science team at Barnes & Noble Education (BNED) yielding their first D2C product; I led the ad inventory monetization group at WarnerMedia which created the methodology for WM's linear ad-household-addressability.

EXTENSIVE AND VARIED PUBLICATION DOSSIER; EFFECTIVE EDUCATOR, PROFESSOR, COLLABORATOR AND COMMUNICATOR WITH PROVEN ABILITY TO LEAD MULTIDISCIPLINARY TEAMS. I have a record of achieving exceptional results via innovative and specialized approaches to business imperatives & research questions. I have led multifunction teams from project inception to completion to viable production for: credit scoring, e-commerce, retail & multi-media corporations. I have collaborated and supervised professionals which has won my teams a proven record of impactful, efficient work.

INNOVATIVE RESEARCHER, PROBABILIST & THEORETICAL METHODOLOGIST WITH A PROVEN ABILITY TO LEAD MULTIDISCIPLINARY TEAMS. I have authored widely cited, novel work in Statistics, Econometrics, Probability & Information Theoretic Measures for Machine Learning & Artificial Intelligence, and Demand for Digital Media Inventories. I serve on: the Data Science Advisory Council at Seton Hall University, the Advisory Board for the Ivan Allen College at the Georgia Institute of Technology, the Advisory Board at Barnes & Noble Education & on the advisory board at Modal Education.

FLUENT IN COMPUTATION: I program chiefly in R, Python, and (Hive/Hadoop/Snowflake) SQL. I am fluent in Maple, Mathematica, Matlab, Neo4j, SAS, SPSS, SQL, C++, Java, Gtk, RGtk, & Visual Basic. Conversant in Julia & Clojure. My teams have worked in AWS/Sagemaker, Azure, Databricks, Dataiku, Condor/HTCondor, etc.

EDUCATION

Postdoc	**STANFORD UNIVERSITY**; Palo Alto, CA Markov Models, Distributions with Fixed Marginals, Copula Models	2008
Postdoc	**DUKE UNIVERSITY**/SAMSI; Durham, NC Multivariate Data, Extreme Valued Data, Bayesian Methods	2007-2008
Ph.D.	**COLUMBIA UNIVERSITY**; NY, NY Probability, Statistics & Environmental Engineering	2008
M.Phil	**COLUMBIA UNIVERSITY**; NY, NY Probability	2007
M.A.	**COLUMBIA UNIVERSITY**; NY, NY Statistics	2002
B.S.	**GEORGIA INSTITUTE OF TECHNOLOGY**; Atlanta, GA Industrial Engineering & Public Policy	2000

SELECTED PROFESSIONAL EXPERIENCE

Gumbel Demand Acceleration, *Head of Science* 2023-present

SaaS (software-as-a-service) company for demand acceleration in digital media for content holders. Pricing and demand modeling for content rights speculation, artist & content-tranche level deal modeling.

- Dynamic metering of Demand Affinity on Segmented Audiences: *Patent Pending*
    - Via a novel process on listening demand data – we've created "Metering Engine" that yields informative metricization of sound and audience features as demand affinities. This "Metering Engine" is packageable as enterprise level software, as a consumer-facing app, within hand-held hardware or as a metadata generating process in a sound curation and delivery pipeline.
- Enhanced lyric transcription via genre and sound awareness: *Patent Pending*
    - By parsing and modeling genre and audience, as well as sound, information we've created a precise and quick method for lyric transcription.

Warner Music Group, *Senior Vice President – Data Science* 2020-2023

Global Music Recording & Publishing company; home to artists, songwriters & musicians across many genres. I lead WMG's Data Science group: a team of Statisticians, Operations Researchers, Psychometricians and Economists & Data Engineers.

- Instantiated Agile Product Process Discipline
    - Instantiated Jira/Roadmunk documented production process. Ported siloed workstreams to cloud-based testing, development & display environments using Databricks, RStudioServer, Looker, Shiny & Tableau. Created stable documentation and code repositories in Git/Jira/Confluence. This created regular procedures, a persistent work repository to refactor workstreams as iterative scientific investigations, and streamlined the R+A group into a modern data science benefit to WMG.
- Created & Productionized Demand Forecast Model
    - My team created and instantiated fully Bayesian, productionized Hierarchical Linear Model to forecast track level music demand in the presence of a suite of features that we derive from aural, temporal, ambient & marketing information. The model yields a sophisticated context for understanding who listens to what, where, when and how. The framework is adjustable, posterior estimation focused, and sensitive to short-term change points; when the data or patterns that affect listening demand change – the framework adjusts flexibly and automatically.
- Created & Productionized Networked Listening/Similarity Models
    - My team created and instantiated a networked (graphical) model for artist/track similarity. The model yields a straightforward and computationally inexpensive way to segment streaming data for precise forecasting. The model topology allows models to be fit via graph traversal on small neighborhoods; the topology itself is learned from multivariate semantic embeddings of aural characteristics and listening demands.
- Created & Productionized Release Scheduling Model
    - My team suffixes expected song streaming/listening demand with nuanced and specific (by territory, publisher, listening segment, etc.) scheduling, marketing & playlisting recommendations using a stochastic programming model that adjusts marketing and promotional recommendations 'on the fly'.

- Created & Productionized Deep Learning IDE for asset augmentation
    - My team placed WMG inventory (tracks, songs, etc.) in an augmented repository. We used deep learning/Neural Network class models to create representations and extract and modify audio stems (vocal, instrumental, lyrical). These metadata power use cases in demand acceleration: finding the best sync opportunities, creating 'music-as-utility' assets, and predicting listening demand from sound. As well, this repository is training data for my team's work in generative AI: creating videos, sounds, & album art from our sounds quickly and scalably.
- Two Patents Pending
    - *Systems, Methods & Media for Transcribing Lyrics*
    - *Dynamic Metering of Demand Affinity on Segmented Audiences*

WarnerMedia, *Director – Data Science* 2019-2020

Linear, Digital & Multimedia company comprised of multiple legacy television networks and motion picture studios as well as digital entertainment and news portals, including: CNN, HBO, TNT, TBS, Bleacher Report, etc. I directed Data Science for the Monetization and Cross-Platform Measurement groups. I led a team of U.S. based Statisticians, Operations Researchers & Operations Analysts tasked with maximizing revenue from ad placement and audience targeting. We worked in R, Python, CPLEX, and in C++. We reported in R/Shiny, PowerBI, Tableau & Looker.

- Instantiated Agile Product Processes
    - Instantiated Jira/Roadmunk documented production process. Ported siloed workstreams to cloud- based testing and development environments using Domino. Created stable documentation and code repositories in Domino/Git. This created regular procedures and a persistent work repository for redundancy.
- Revised Forecasting Model Suite
    - Replaced Random Effects model for schedule forecasts (Linear/TV for Turner Networks – TNT, TBS, CNN, Etc.; Digital – Bleacher Report, NBA Digital, CNN Digital) with Hierarchical/Multi-level Bayesian, time-dependent & stochastic process predictive models. This innovation improved forecast precision (Mean Average Percent Error – MAPE) by 30-40% with an estimated incremental yield to business on order of $100M yearly revenue.
- Revised Optimization Model Suite
    - I replaced LP/MIP class of discrete programming models with 2-stage LP and fully stochastic programming models. This innovation reduced scheduling error, reduced *ad hoc* schedule manipulation – raising FTE equivalent efficiency. This yields an estimated incremental impact on the order of $10Ms – yearly – for a business unit on order of $100Ms annual revenue.

Barnes & Noble Education, *Head - U.S. Data Science* 2018-2019

Education retail, services, publishing & technology company that provides – *inter alia* – predictive services to undergraduate/graduate universities using Student Information System (SIS) and Learning Management System (LMS) data. I led data science initiatives in the United States and managed a team of outsourced data scientists based in Mumbai, India. We worked in **R**, Python, Elastic Search, Kibana and Javascript.

- Instantiated Standardized `Tear Sheet` Reporting Format.
    - I designed in-term and end-of-term reporting schema for model accuracy and discovery rates.
    - I created an automated process to ingest discovery rate data and compile LaTeX to PDF documents which illustrate model accuracy and discovery rates.
    - Streamlined reporting process and reduced utilization of data scientists on reporting tasks by half.

- Created text mining methodology for course title / textbook mapping.
    - I created mappings between courses titles and subject area textbook information (from publishers) using Naïve Bayes classifications on syntactic tokenizations of course title and textbook subject corpuses.
    - Standardized *ad hoc* procedure for OER resource creation
- Created Enterprise Level Data-Lake
    - I created, using Elastic Search, a repository for SIS/LMS data *across* university entities with unique, anonymized, GDPR compliant indexing
    - Instantiated a stable code-base for regularized data ingest, model training, and model fitting: functionalized, portable, & able to ingest, identify and map exogenous predictors.

Dun & Bradstreet, *Sr. Data Scientist*

2016-2018

Business services company that provides commercial data to businesses on credit history, business-to-business sales and marketing, counterparty risk exposure, and social identity matching. I reported to the CEO's office in a direct line from the Chief Data Scientist. I worked with teams of diverse specialists on project bases. Here my team and I worked in: R, Python, Hive/Hadoop, SQL, SAS & Mathematica.

- Confounding Characteristics of Language in Unstructured Data.
    - I designed research and led a team to quantify the presence of semantic ambiguities in short and long form unstructured data.
    - I developed a model for quantification of confoundedness via Grammar & Usage, Neologism & Sarcasm in English natural language.
    - I augmented Python based toolkits and created a Markovian Process Model to quantify and classify these characteristics.
    - I led a team in the creation of a Mathematica based tool – with a Python back end – to score unstructured text.
- Briefed NSTAC (National Security Telecommunications Advisory Committee): Big Data Analytics/Emerging Technologies, 10.18.2016
    - Delivered prospective brief on role of Quantum Computing and Algorithms in credit scoring and anomaly discovery in business analytics.
- Multivariate Hidden Markov Process for Match/Inquiry Assessment.
    - I created a Markov Process model to assess abnormality in mappings from a discrete `grading scheme' to ordinal goodness measures for validation of customer inquiries.
    - I directed a team in the instantiation of the model to a cloud-based tool.
- Patent Pending for Discovery of Malfeasant Actors via Networked/Graphical Models.
    - I created a model-diagnostic based method to segregate malfeasant actors in predictive models for behavior (payment latency) on graphical structures for networked data
- Multivariate Models for Machine Learning:
    - I instantiated an AWS (Amazon Web Services)/R/Elastic-Search environment to perform unsupervised classification and identification of anomalies in business data.
- Patent Pending for Demand Based Marketing Recommendation Engine
    - I created a methodology, via exploitation of customer inquiry data, to classify and recommend prospective future customers. This is a recommendation engine embedded in a clustered network data model.
- Briefed NSTAC (National Security Telecommunications Advisory Committee): Cybersecurity/Emerging Technologies, 8.15.2017

- I prepared prospective brief on role of Quantum Computing and Algorithms in response to DDoS & Cryptographical attacks.

SELECTED PUBLICATIONS

How & Why to Use Audience Segmentation to Maximize (Listener) Demand Across A Digital (Music) Portfolio
2024
MAA 2024

- A First Principles derivation of distribution of listening demand and representation as utility curves
- A Bayesian workflow for effect estimation and demand maximization

Open Set Recognition for Music Genre Classification
2022
ISMIR 2022

- An algorithm for music genre classification
- Explores boundary detection, via first principles methods on type III EVD (extreme-valued distributions), to quantify sound content & audience segment set origination.

Methodologies and Model for the Detection of DDoS Attacks on Cloud Computing Environment
2016
Advances in Intelligent Systems and Computing

- A hybrid statistical and probabilistic method to detect DDOS attacks.
- Developed novel statistics, based on probabilistic dependency measures, to quantify distributions for multivariate feature dependence.

Statistics for Re-Identification in Network Models 2015
Network Links: Network Analysis

- A 1st Principles Probabilistic derivation of statistics for re-identification on Network/Graph Topologies
- Derived network similarity scores for graph models of various generative types: Barabási-Albert, Random, etc.
- Yielded hypothesis testing statistics for similar entity behavior in network models, i.e. under relational dependency.

Monitoring the UNDP Millennium Goals 2013
Social Indicators Research

- Generalized Bayesian Methodology for Indexes – the UNDP MDGs are the special case.
- Introduced a unique, Probabilistic technique for constructing administrative indexes which yields ranked lists with confidence intervals.
- Modeled health, infant mortality, morbidity and fecundity outcomes for the MDG countries.

Using Lorenz Curves to Examine ITQ Consolidation in New Zealand Commercial Fishing 2013
Marine Resource Economics

- The Lorenz Curve – a homeomorphism of the Cumulative Distribution Functions, and an inverse mapping of the Gini Coefficient – can be partitioned across discrete groups using the theory of differential equations. The method is derived and illustrated.
- Demonstrated the presence of consolidation in fishing rights over a 20 year period in New Zealand, indicating the presence of market monopolization.

Copula Based Multi-State Hazard Model: 2011
An Inferential Methodology for the Innocence Project:
Proceedings of the American Statistical Association

- Extended the Multi-State Hazard Model – a generalization of Survival Analysis – to allow flexible conditional dependency between states.
- Used the Copula equivalence for the Chapman-Kolmogorov equations to parameterize conditional dependence among states, which yields system sojourn time and probability.
- Demonstrated best practices for Innocence Network lawyers and interns with respect to case intake, retention and flow procedures.

Diagnostics for Multivariate Imputation 2008
Journal of the Royal Statistics Society

- Created a methodology to assess and test the propriety of imputed data, at the time an open question in data analysis. The methodology has been implemented in **R** and SAS.
- Illustrated the technique on the 2002 Environmental Sustainability Index (ESI).

El Niño & Drought in Southern Africa 2003
The Lancet

- Used GIS data to classify drought via percentage of precipitation vs. administrative declaration.
- Probit/Logit linear models with drought classification and functions on Sea Surface Temperatures (SST) demonstrated co-dependency among the El Niño phenomenon and drought classification: most notably in Southern Africa.
- Demonstrated the effect of El Niño on drought incidence, health status & morbidity of populations.

SELECTED ACADEMIC EXPERIENCE

Adjunct Professor of Statistics 2016-present
SETON HALL UNIVERSITY, SOUTH ORANGE, NJ, USA

- **Statistics for Biologists, Mathematical Statistics for Data Scientists, Mathematical Statistics, Deterministic Operations Research for Data Science.** Suite of classes from intermediate to graduate level with an **R** based curriculum, the first in the department. **Teacher of the Year in 2017**.

Visiting Professor of Probability & Statistics
2014-2016
UNIVERSIDAD DE CUENCA, CUENCA, ECUADOR

- **Designed first `Postgrado' class in Algorithms, Data & Computation** for Facultad de Ingenería.
- **Designed new program in Applied Mathematics** and new classes in Statistics, Probability and Optimization.

SUNY Faculty Fellow Professor of Mathematics & Environmental Science 2012-2015
BINGHAMTON UNIVERSITY, BINGHAMTON, NY, USA

Assistant Professor of Statistics & Industrial Engineering 2008-2012
GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GA, USA

- Designed Graduate course in Theoretical & Applied Dependency. Consistently had highest teaching marks among Statistics faculty.
- Authored/Co-Authored eighteen published journal papers in three years.

SERVICE, HONORS & AWARDS

MAA Chan-Stanek Lecture	2024
Distinguished Alumni of the Year, Georgia Institute of Technology	2021
Board Member, Modal Learning	2022-Present
Board Member, Barnes & Noble Education	2021-Present
Board Member, Georgia Institute of Technology: Ivan Allen College	2021-Present
Panelist, Division of Mathematical Sciences, National Science Foundation	Winter 2012
Member, Georgia Tech Institutional Review Board	2010-2012
Session Chair, INFORMS	2007, 2008, 2010, 2011
Faculty Council Chair, Barnes & Noble Education	2018-2019
Data Science Advisory Council, Seton Hall University	2018-Present
Class of 1969 Teaching Fellow, Georgia Tech	2011-2012
GT-FIRE Program to Support Innovative Research	2010-2011
Young Practitioner INFORMS	2009
VIGRE Fellow, Stanford University Statistics Department	2008
SAMSI Postdoctoral Fellow, Statistical and Applied Mathematical Sciences Institute	2007-2008
CFD Predoctoral Fellow, Haverford College	2006-2007
NSF IGERT in the Mathematical and Earth Sciences.	2003-2007
Best Algorithm, CAARMS	2004