Brandon **Dey**

Sr. Data Scientist & Machine Learning Engineer

912 Channelside Drive Unit 2311, Tampa, FL 33602

(503) 866 - 8354 | brandonwdey@gmail.com | brandondey.github.io | brandondey | brandondey

Summary

I'm a creative data scientist with 8+ years experience using machine learning, applied statistics, and AI techniques for companies of all stripes. I thrive in complexity, am motivated by endless curiosity, and take pride in my hard-won ability to communicate with people at all levels of seniority and technical acumen. With expertise in end-to-end data analytics processes, from ETL to executive level presentations, I've coupled chops in **R, Python, SQL, AWS, Git, and Hadoop** with a theoretical understanding of applied mathematics and machine learning to **design experiments, test hypotheses, mine insights, and build data-driven, production-grade systems** that make the most of available data.

I science on System76's **Thelio Major** for scientific computing. **RAM:** 64GB; **Hard drive:** 1TB; **CPU:** Ryzen Threadripper 2920X (12-core/24-thread @ 4.3 GHz); **GPU:** Radeon RX 590 with 2304 stream processors.

Education

University of Minnesota — *Minneapolis, MN*
Applied Statistics, GPA 4.0/4.0 — *2013*

University of Wisconsin - Milwaukee — *Milwaukee, WI*
BA in Economics, GPA 4.0/4.0 — *2015*

Data Camp Certifications — *2018*

- Statistical Modeling I, II
- ggplot2 I, II, III
- Machine Learning in R
- Reporting with R Markdown
- Correlation and Regression
- + a half-dozen more

Cloud Certifications — *Forthcoming*

- AWS Cloud Practitioner
- Azure Fundamentals
- Azure AI Engineer Associate
- Azure Data Scientist Associate
- Azure Data Engineer Associate

Programming Skills

Languages

- R (Highly adept, 8 years) {tidyverse, stats, MASS, custom functions}
- Python (Proficient, 4 years) {tensorflow, keras, Beautifulsoup4, numpy, scikit-learn, pandas}
- SQL (Highly adept, 8 years)
- SPSS (Proficient, 2 years)
- LaTeX (Proficient, 3 years)
- STATA (Proficient, 1 year)
- Scala (Working knowledge, <1 year)

Technologies

- Microsoft SQL Server (Highly adept, 8 years) {SSMS, SSRS, T-SQL}
- MySQL (Highly adept, 5 years) {Workbench}
- Microsoft Visio & LucidCharts (Proficient 5 + years)
- R Markdown (Experienced, 4+ years)
- Git (Lab & Hub) (Adept, 3+ year)
- OpenStreetMap (Proficient, 1+ year)
- Hadoop/Hive/Impala (Proficient, 2+ years)
- Tableau (Working knowledge, < 1 year)
- R Shiny (Proficient, 2 years)
- AWS (Proficient) {SageMaker, Rekognition, Redshift}
- Databricks (Proficient, <1 year)
- ThoughtSpot (Proficient, <1 year)
- Snap Logic (Proficient, <1 year)

Work Experience

Wavicle Data Solutions — *Remote/Chicago/Tampa*
Chief Data Scientist, Emerging Technologies Practice — *April 202 - Present*

- This role is 70:30 technical:managerial. As the primary data science subject matter expert of this new practice, I'm responsible for establishing the firm's technical credibility by building and open-sourcing prototypical machine learning systems using emerging technologies including, Databricks, ThoughtSpot, and SnapLogic.
- As the third member of this new practice, I help set and implement technical strategy, screen and hire new teammates (solution architects), and present technical solutions in client meetings, spanning grocery, retail, oilfield transportation, eCommerce, public organizations (municipaliies and port authorities, and manufacturing (medical and product)
- Currently building: an advanced music search engine that finds songs better. It can be asked to find songs based on adjectives (red, jumpy, destitute), feelings (aggressive, relaxed, happy, etc.), cultural/topical/thematic references (about Eminem but not by Eminem), acoustic characteristics (key, chords key, tone, rhythm, pitch, timbre), lyrical qualities (rhyme schemes, degree of profanity, etc.), and more–all at scale thanks to Databricks and Apache Spark. Specifically, I use the Databricks Unified Data Analytics Platform (Apache Spark, Delta Lake, and MLflow) to ingest, prepare, and store music metadata from across the web, shoehorn it into a Natural Language Processing pipeline (LDA topic modeling, sentiment analysis, latent semantic analysis), and wrap it in an R Shiny search interface, to enable music lovers to find songs with impressive precision by leveraging a data-driven search strategy based on the same mechanics that power Spotify.

Publix — *Tampa, FL*
Senior Data Scientist, Digital Machine Learning Group, IT — *May 2020 - May 2021*

- This is a product role in which I build production-grade machine learning models to scale across 1,300+ grocery stores.
- Engineered Publix's first product recommendation engine for online, in-store pickup orders, achieving 15 percent F1 score. Version 1.0 was standard a item-based collaborative filtering algorithm with custom features that enabled the business to control recommendation characteristics, like product diversity, frequency, and how customer purchase history is weighted. The algorithm pays more attention to customers with shallow purchase histories than it pays to those with deep histories in effort to capitalize on targeted shopping behavior.
- Version 2.0 of the production recommendation engine captures temporal dynamics of ever-evolving consumer preferences and drifting item bias. Parameters are fit via a custom-coded stochastic gradient descent optimization framework inspired by The BellKor Solution to the Netflix Grand Prize paper.

The Kroger Company (84.51, Data Science Arm of Kroger) *Portland, OR*
Senior Data Scientist, Operations Research *March 2019 - April 2020*

- Team lead for measuring the effect of our proprietary stacked ensemble that predicts how likely a store clerk will churn in the short term. I architected and engineered an automated evaluation tool that operates within the quasi experimental context of our interventions. Under the hood, it's a network of Bayesian structural time series models, not unlike Google's CausalImpact package, but with greater transparency and improved flexibility for more precise selection of the synthetic control group. In production.
- The aforementioned HR model includes legally protected classes, so statistical equity is paramount. I built an automata that detects and addresses any bias in our machine by evaluating true/false positive rates across class groups and applying a novel post hoc weighting scheme when needed. In production.
- Responsible for engineering 15 features in our HR model with 100+ others, for three of which I used anomaly detection (anomalize) to root out more signal. 4-5 of my features land in the top 10 most predictive features every month. In production.
- Designed and engineered a suite of unit tests for an critical ETL pipeline that robustly checks various consistencies between databases and emails the team an HTML summary. In production.
- Lead team in software engineering best practices by functioning as gatekeeper who reviews all pull requests before committing to master/production. Spearhead meetings to keep team accountable to strong software practices.
- Improved forecasts of Pickup demand by adding engineered features for local weather conditions, resulting in more accurate estimates of how much labor a store ought to schedule, which enabled our optimization engine to create even better labor schedules for Pickup.

Fisher Investments *Portland, OR*
Team Lead Data Science, Global Marketing *Oct. 2018 - March 2019*

- Researched, designed, coded, and pitched a novel optimization solution to Global Marketing's multi-channel marketing problem space, which was a chain of channel-level Adstock models whose output was fed into a constrained optimization model. The solution allocated marketing spend to the most productive channels, while balancing cost and quality of leads.
- Increased key sales quality metric by 9 percentage points while maintaining sales volume with two random forest models that discovered new predictive features from online retirement calculator which legacy solution missed. Solution gave marketing management control over cost and quality of sales leads for first time.
- Increased accuracy of propensity mail model by one percentage point for "free" by using active learning when downsampling non-responders who outnumbered responders 20 to 1.
- Optimized paramaters of legacy elastic net lead scoring model for remarketing mail campaigns, resulting in more precise targeting.

Direct Supply, Inc. *Milwaukee, WI*
Data Scientist, Data Science *March 2017 - Sept 2018*

- Served as technical contributor and editor of an internal data science journal focused on (i) discovering actionable customer insights, (ii) recommending what to do about them, (iii) unpacking how to monetize the data generally and (iv) demonstrating the power of machine learning, written for an executive audience.
- Identified an opportunity to create a new line of business after using a Random Forest decision tree to predict if a customer turnover would generate a net profit or loss.
- Designed, built, and evaluated a system that generates recurring monthly call and email campaigns by targeting customers most likely to buy in timeframe using logistic regression. Achieved 13 percentage point boost in clicks and 5 percentage point increase in purchase rate over a 20% holdout group in first 12 months in production (R::stats::glm).
- Built a Latent Dirichlet Allocation Topic Model that predicted what free-form text-based work orders were about, allowing engineering to trigger real time advertising on topical key words entered by mobile app users (R::topicmodels).
- Estimated causal effect of a new product line on existing margin using bayesian structural timeseries models (in absence of randomized experiment), enabling product managers to proceed to rollout more product (R::CausalImpact).
- Predicted date range a skilled nursing facility was to be inspected with 85% accuracy using Stratified Cox survival model to monetize publicly available data from the Centers for Medicare & Medicaid Services (CMS) (R::survival).
- Automated the analysis of a biannual Net Promoter Score survey in R (was living in SPSS), reducing 280 hours of work from 1.5 FTE's to 14 hours of work by 1 FTE using custom-developed library of R functions.
- Helped eCommerce executive leadership meet their KPI of retaining, growing, and finding customers by visualizing a time series of customer lifecycle (new, full, lapsed)(R::TraMineR).
- Estimated opportunity cost of discontinuing support of Internet Explorer 8 to help IT assess impact.
- Partnered with key internal stakeholders for the development of scalable machine learning and predictive modeling solutions. Provided analytical consultation to partners across all departments, ranging from executive level to team members, using R and SQL.
- Created dashboards, reports, and analyses that summarized complex findings to help internal partners make smarter decisions.
- Developed high level data driven insights that confirm or refute partner's hypothesis. Responded to complex inquiries with actionable insights and advice.
- Defined analytical and reporting requirements with partners.

Direct Supply, Inc. *Milwaukee, WI*
Business Intelligence Analyst, Data Science *Dec. 2015 - March 2017*

- Authored an Abandonded Shopping Cart Report that informed sellers when customers in territory placed product in their shopping cart but hadn't purchased yet, which led to a 4 percentage point lift in sales.
- Authored complex SQL queries/views/functions and reports to meet key KPIs, which varied across teams, departments, and business units.
- Acted as an advisor to stakeholders by having knowledge of data structures and system capabilities.

Direct Supply, Inc. *Milwaukee, WI*
Reporting Analyst Intern, Data Science *Oct. 2014 - Dec. 2015*

- Fulfilled reporting requests by building and updating reports and dashboards using SQL, SSRS, and Microsoft Visual Studio.

University of Wisconsin - Milwaukee *Milwaukee, WI*
Economics Research Assistant, Undergraduate Research Fellowship *Sept. 2014 - Dec. 2014*

- Performed tasks to advance the research project of addressing whether a 5 cent plastic bag tax in Washington D.C. led to a decrease in food consumption, employing a difference-in-difference model, under the guidance of Dr. Laura Grant.

Literacy Services of Wisconsin *Milwaukee, WI*
Math Basics Coordinator, Intern *May 2014 - Aug. 2014*

- Developed a new curriculum of basic math courses for adult students and taught all math lessons 1:1.

Side Projects

Natural Language Processing

Dicto – The smart dictionary *Portland, OR*
Forthcoming

- Imagine a logophilic bot that fetches the precise word for the precise context. Search for words based on intuitive parameters like "words that sort of mean [adjective] but are less like [synonym]", word rareness, modernness, domain relevance (legalese, medical terminology, sports vernacular, male(female)ness, etc.), the word(s)' "[adjective]-iness", tensile strength, etc. It also displays each word's association to contentious topics like race, slavery, gender, and other extremist topics so the user is informed of – and can sidestep – the word's charge in the American zeitgeist.
- Trained on data from: **Google nGram**, **Merriam-Webster's Collegiate® Thesaurus**, and **DataMuse.**

Pop: Palindromic Prose, Pronto *Portland, OR*
July 2019

- I designed and engineered a novel algorithm that creates palindromic prose from the word sequences of inputed corpora. Powered by a custom variant of the **Rabin-Karp Algorithm** and **Apache OpenNLP**.
- There's an accompanying D3.js viz that visualizes the intuition of how it works, similiar to **this**.

Text Mining 3 Years of My Family's Text Thread *Milwaukee, WI*
March 2015

- Created personalized visualizations for 10 family members who exchanged texts with one another over a three year period.
- Insights included who was most likely to initiate conversation, who was most exclamatory(!!!), most inquisitive(???), most emoji-tional, night owls, early birds, topical trends over time per person, simple word clouds, and predicted response time given the initiating interlocuter.

Writing

Machine Learning, Applied Statistics, Data Science *Cambridge, MA*
AI & Data Science Journalist, www.OpenDataScience.com *MAY 2018 - PRESENT*

- A Survey of Popular Ensembling Techniques – Part 1
- From Idea to Insight: Using Bayesian Hierarchical Models to Predict Game Outcomes Part 2
- From Idea to Insight: Using Bayesian Hierarchical Models to Predict Game Outcomes Part 1
- Active Learning: Your Model's New Personal Trainer
- ODSC West 2018 Review
- An Introduction to Sentence-Level Sentiment Analysis with sentimentr
- Where's Humanity's Algorithm to Solve Poverty?
- A Beer Lover's BFF? An Introduction to Geospatial Interpolation via Inverse Distance Weighting
- Comparing Five Different Smooths – Which One Rules Them All?
- A Short Summary of Smoothing Algorithms

Honors & Awards

2014 **Formal Commendation**, From English Chair for "outstanding writing and clarity of thought in ENG 350"
2012 **Scholar's Award (Best exam score)**, Delta Sigma Pi
2010 **Most Likely to Succeed**, Coleman High School
2010 **English Award**, Coleman High School