Chase Joyner

chasejoyner77@gmail.com | 704-830-4730 | Charlotte, NC | www.chasejoyner.com

Highlights

Quantitative skills: Machine Learning, Random Forest, Boosting, Neural Networks, Natural Language Processing, Large Language Models, Generative AI, Transformers, Classification, Regression, Penalized Regression, Timeseries, Markov chain Monte Carlo, etc.

Software: Python, PySpark, PyTorch, R, SQL, GitHub, LaTeX, Databricks, AWS, Delta Lake, SAS, Excel, etc.

Professional summary: Model development, experimentation design, data analysis, and presenting results to senior leaders, business partners, and model validators with a track record of delivering successful outcomes.

Education

Ph.D. in Statistics, Department of Mathematical Sciences, Clemson University — Dec 2019

M.S. in Statistics, Department of Mathematical Sciences, Clemson University — May 2016

B.S. in Mathematics, Department of Mathematical Sciences, Clemson University — May 2014

Experience

Data Scientist, Teladoc Health — Charlotte, NC
Senior Data Scientist — Feb 2024 – Present
Data Scientist III — Apr 2022 – Feb 2024

- ◇ Developing a content generation LLM using Azure's OpenAI REST API to personalize marketing communications based on internal member data. The model adheres to business constraints and allows marketing to tailor more than 25 different communications to identified member categories. This replaces a single generic communication and increases member engagement through relatable content.
- ◇ Developed a tree-based recommender model informed by internal and external data (CDC and epidemiological studies) which resulted in over 42% of the member base being targeted by marketing outreaches. This model has also been integrated in other areas of the business and won the Teladoc Health Innovation Award for its proven impact on the business.
- ◇ Developed an XGBoost model to score members based on their propensity to engage with the expert medical opinion service. The model was built on historical campaign data which resulted in more than 50% lift in conversion rates when targeting top deciles, aligning with the company's campaign optimization efforts in improving ROI metrics. The model is used for each quarterly expert medical opinion campaign.
- ◇ Developed machine learning propensity models to continually score members based on their likelihood to respond to marketing channels or engage in certain products. Model performance is measured through outcome driven metrics such as accuracy, recall, or precision and is monitored quarterly to guide timely model retuning.
- ◇ Developed a Markov chain expected return time framework to guide timely communications for members who deviate from their normal account activity. This replaced a fixed 10-day inactivity flag with a model-based approach that dynamically estimates a member's normal patterns, lowering the risk of member churn.
- ◇ Built and conducted statistically sound experimental designs to gather member behavioral data. This insight is useful for recommending member personalization and engagement initiatives to marketing partners in an effort to increase member engagement and return on investment (ROI).
- ◇ Contributed to and maintained a large scale data science GitHub repository which houses a suite of registered models, database connections, and databricks notebooks used organization wide.
- ◇ Software: Python, PySpark, Databricks, GitHub, AWS, Delta Lake, SQL, etc.
- ◇ Techniques: Decision trees, random forests, XGBoost, Markov chains, experimental designs, A/B testing, etc.

Quantitative Operations Associate, Bank of America Charlotte, NC
Assistant Vice President Sep 2021 – Apr 2022

- ◇ Developed quantitative volume forecasting models for the Bank's capacity and strategic workforce planning in the First Mortgage and Home Equity loans space.
- ◇ Led the automation and standardization of volume forecast modeling to be used organization wide, which reduced model development time by over 50% and eliminated error-prone manual processing.
- ◇ Collaborated with business leaders to enhance the forecasting of volumes based on historical trends and changes in business processes.
- ◇ Presented quarterly volume forecasts to business executives and finance partners for approval.
- ◇ Software: Python, SAS, GitHub, Excel, etc.
- ◇ Techniques: ARIMA, ARIMAX, X13, etc.

Quantitative Finance Analyst, Bank of America Charlotte, NC
Assistant Vice President Oct 2019 – Sep 2021

- ◇ Summary
  - Developed statistical models on complex data sets for the Bank's wholesale portfolio. Modeling experience covers loss forecasting for regulatory CCAR, CECL, and IFRS 9 reporting.
  - Organized, documented, and presented results to senior leaders, business partners, and model validators.
  - Software: Python, Linux, GitHub, SQL, Excel, etc.
  - Techniques: Cox proportional hazard rate model, maximum likelihood estimation, logistic regression, likelihood ratio test, variable selection, bootstrapping, goodness-of-fit metrics, etc.
- ◇ Wholesale Loss Forecasting Modeling (CCAR, CECL, IFRS 9)
  - Led the redevelopment of the defaults and ratings transition model, one of the largest and high-risk models for the Bank's commercial and industrial portfolio. The model outputs point-in-time quarterly transition matrices used for CCAR, CECL, and IFRS 9 reporting.
  - Developed, implemented, and documented various sensitivity tests to assess potential operational risks in the modeling framework. Testing involved derivation and implementation of the least absolute shrinkage and selection operator (LASSO) for variable selection and incorporating constrained maximum-likelihood estimation to increase model output granularity for more precise modeling.
  - Contributed to a large scale collaborative model code repository using version control software and wrote unit tests to ensure code changes adhere to expected output. This code base is used across the Bank's risk modeling organization.
  - Improved model accuracy and flexibility by incorporating industry specific macrovariables not previously considered in the modeling framework. This was a step forward in the Bank's efforts to climate risk modeling and allowing more accurate stress forecasts as industries are affected differently during stress.
  - Performed data assessment and variable selection using likelihood ratio test, Wald test, and backward stagewise selection. Model explanatory power was demonstrated through default, downgrade, and upgrade rate backtesting and various goodness-of-fit metrics such as AUC and RMSE.
  - Presented model forecasts to business stakeholders and communicated material model updates that drove changes in expected credit losses.
- ◇ Wholesale Capital Reserve Modeling
  - Accepted model hand-off halfway through model redevelopment for the probability of default (PD) component. The model is developed to be compliant with the Third Basel Accord (Basel III) for the Bank's risk weighted asset (RWA) and economic capital (EC) planning.
  - Documented the PD model's redevelopment and communicated the updates to business stakeholders and model validators for a successful validation.
  - Performed COVID-19 impact analysis to understand model implications and to defend the model was fit for use during the pandemic.

Department of Mathematical Sciences, Clemson University Clemson, SC
Graduate and Research Assistant Aug 2014 – Dec 2019

- ◇ Developed a mixed effects logistic regression model for predicting maize stalk lodging by accounting for genetic similarity, environmental factors, and predicting intermediate phenotypic values, such as maize stalk strength. Successfully showed that stalk strength is more effective than existing phenotypes. Research sponsored by National Science Foundation.
- ◇ Developed a Bayesian mixed effects logistic regression model with variable selection to analyze data from any group testing algorithm while accounting for imperfect testing. Improved existing methods by including in the model which clinic each patient visited. Research sponsored by National Institute of Health.
- ◇ Analyzed rice data provided from fields in Indonesia to develop a mixed effects linear regression model accounting for complex genetic similarity. Improved techniques for genome-wide association studies to perform efficient joint analyses and found several genes linked to yield count. Research sponsored by Biorealm Principal Investigator.
- ◇ Developed univariate and multivariate Bayesian models to estimate the optimal biomarker density threshold in pooled testing of individuals for various diseases.
- ◇ Onboarded and mentored new students to research projects and introduced them to the Clemson Palmetto Cluster, a Linux based supercomputer.
- ◇ Taught various business calculus and statistics courses to students at Clemson and Indonesia.
- ◇ Software used: R, Linux, Clemson Palmetto Cluster, Excel, etc.
- ◇ Techniques used: Logistic regression, generalized linear models, nonparametric regression, mixed effects models, Bayesian statistics, spike and slab priors, Markov chain Monte Carlo (MCMC) methods, expectation-maximization algorithm, etc.

Publications

- **Joyner, C.**, McMahan, C., Tebbs, J., and Bilder, C. (2024+). A multivariate Bayesian mixed effects model with variable selection for multiplex group testing data. In preparation.
- Yusuf, I., Miskad, U., Lusikooy, R., Arsyad, A., Irwan, A., Mathew, G., Suriapranata, I., Kusuma, R., Pardamean, B., Kacamarga, M.,Budiarto, A., Cenggoro, T., Pardamean, C., McMahan, C., **Joyner, C.**, and Baurley, J. (2024+). Genetic risk factors for colorectal cancer in multiethnic Indonesians. *Scientific Reports*, 11, 9988.
- **Joyner, C.**, McMahan, C., Tebbs, J., and Bilder, C. (2020). From mixed-effects modeling to spike and slab variable selection: A Bayesian regression model for group testing data. *Biometrics*, 76, 913-923.
- Sekhon, R., **Joyner, C.**, Ackerman, A., McMahan, C., Cook, D., and Robertson, D. (2020). Stalk bending strength is strongly associated with maize stalk lodging incidence across multiple environments. *Field Crops Research*, 249.
- **Joyner, C.**, McMahan, C., Baurley, J., and Pardamean, B. (2020). A two-phase Bayesian methodology for the analysis of binary phenotypes in genome-wide association studies. *Biometrical Journal*, 62, 191-201.
- McMahan, C., Baurley, J., Bridges, W., **Joyner, C.**, Fitra Kacamarga, M., Lund, R., Pardamean, C., and Pardamean, B. (2017). A Bayesian hierarchical model for identifying significant polygenic effects while controlling for confounding and repeated measures. *Statistical Applications in Genetics and Molecular Biology*, 16, 407-419.

Presentations

- Development of multi-scale ideotypes for lodging resistance in maize and sorghum. Apr 2019
  Clemson University; Clemson, SC
- From mixed effects modeling to spike and slab variable selection: A Bayesian regression model for group testing data. *Clemson University; Clemson, SC* Apr 2018
- A mixed effects model for group testing data with variable selection. Mar 2018
  ENAR Spring Meeting; Atlanta, GA

- Bayesian mixed effects model with variable selection for group testing data. Mar 2018
  University of Georgia; Athens, GA

- Assessing the relationship between SNPs and yield in various rice varieties. Nov 2016
  Jakarta, Indonesia

- Bayesian approach of biomarker density estimation using pooled data. Feb 2016
  Clemson University; Clemson, SC

Awards

- Teladoc Health Innovation & Impact Award 2024
- Drennon/Gabet Endowed Memorial Engagement Award 2019
- National Science Foundation Research Traineeship (NRT) Fellow 2018
- Outstanding M.S. Student Award 2015, 2016

Updated: April 2024