

DANIEL K. BERRY

danielberryconsulting@gmail.com · [linkedin.com/in/danielkberry](https://www.linkedin.com/in/danielkberry) · danielberry.io

Education

2015 - 2016	M.S. Statistics Virginia Polytechnic Institute and State University (Virginia Tech), GPA: 3.9/4.0
2014–2015	M.S. Applied Statistics University of Alabama, GPA: 4.0/4.0
2011–2015	B.S. Mathematics (minor in Computer Science) University of Alabama, GPA: 3.5/4.0

Professional Experience

2017 - Current	Data Scientist Allstate Insurance Company Analyzed data and built predictive models using a variety of data sources from embedded devices and smartphones for vehicle telematics applications. Implemented a system in Apache Spark for data ETL and user-specific predictive model building, enabling Allstate to build a different predictive model (GBM/Random Forest) for each of hundreds of thousands of users in minutes.
2015 - 2016	Lead Statistical Collaborator Laboratory for Interdisciplinary Statistical Analysis (LISA), Virginia Tech Provided statistical consulting services for researchers at Virginia Tech including data manipulation (data transformation and statistical programming), data visualization, and data analysis (model building and interpretation). Lead a team of associate collaborators in implementing solutions. Consulted for clients in a variety of departments including business, engineering, architecture, biomaterials, genetics, agriculture, and osteopathic medicine.
2015-2016	Data Science Intern Allstate Insurance Company
Summer 2016	Developed a model to detect car crashes using data from vehicle embedded devices. Aggregated data from a variety of sources and resolutions such as accelerometer readings, GPS trails, and vehicle OBD2 port data using Python, Apache Spark, and Hive. Extracted features using a variety of Python toolkits (e.g. Numpy, Scipy, Pandas, Statsmodels). Trained gradient boosted decision tree ensemble (GBM) models using XGBoost and scikit-learn. Demonstrated 50-fold accuracy improvement over original crash detection logic.
Summer 2015	Developed a GBM model to predict losses from auto accidents using R. Prototyped a natural language feature extraction system in Python to mine information from the notes of accident claims. Demonstrated utility of system by comparing to existing models and showing increased predictive accuracy. Assessed the utility of a business metric for quantifying agent success through visualizations and a Generalized Linear Model based predictive model (implemented in R). Presented results to management to inform future decisions.

Technical Skills

Tools	R, Python, SQL, Hadoop/HDFS, Hive, Apache Spark, MATLAB, SAS, and \LaTeX
Statistical Models	Generalized Linear Models, Generalized Linear Mixed Models, Decision Tree Ensembles (Gradient Boosting and Random Forests), and unsupervised methods (clustering).

Publications

2015	D. Berry , J. Bell, E. Sazonov, "Detection of cigarette smoke inhalations from respiratory signals using decision tree ensembles", in <i>Proc. 2015 IEEE SoutheastCon</i> , Ft. Lauderdale, FL
2015	D. Berry , E. Sazonov, "Clustering technical documents by stylistic features for authorship analysis", in <i>Proc. 2015 IEEE SoutheastCon</i> , Ft. Lauderdale, FL

Community Involvement

2011	Eagle Scout , Boy Scouts of America
------	--