

MICHAEL INGRAM

(636) 484-0824 · micha.j.ingram@gmail.com · <https://github.com/mjingham>
<https://www.linkedin.com/in/michael-ingram-150a37177/>

Data analyst with experience in machine learning, statistical modeling, and consulting. Experienced programmer with knowledge of SQL databases as well as the full data pipeline of data collection, cleaning, analyzing, reporting and visualizations.

EDUCATION

MASTER OF SCIENCE IN STATISTICS, UNIVERSITY OF COLORADO DENVER

Master's Thesis: Classifying Fox and CNN political articles using Random Forest and Natural Language Processing

BACHELOR OF SCIENCE IN STATISTICS, COLORADO STATE UNIVERSITY

PROGRAMMING

- R
- Python
- SQL Databases
- Tableau
- Excel, Power BI
- SAS

EXPERIENCE

Geospatial Data Analyst: T Carta Marine, Denver, CO

August 2020-January 2021

- Space based laser (ICESat-2) data acquisition of shallow water bathymetry, 3-d point cloud editing of satellite derived bathymetric surfaces, development of uncertainty for SBL and satellite derived bathymetric surfaces.
- Marketing analytics with Google Analytics including natural language processing work on emails for ad words.
- Writing python programs for automation of space-based laser data collection workflow.

Master's Thesis: Classifying Fox and CNN political articles using Random Forest and Natural Language Processing

- Ten political articles from both CNN and Fox News were web scraped each day for a month. Articles were cleaned using natural language processing techniques.
- Random Forest and Logistic Regression used to classify articles to their source using the article text and title. Random forest algorithm resulted in feature importance ratings to the top words used in the classification.

Publication: *Factors affecting detection of bimodal sour-savory mixture and inter individual umami taste perception* in the Journal: Food Quality and Preference

- Consulting work for the Denver Museum of Nature and Science where researchers were investigating if the perception of umami changes when acidity is added.
- Data for this experiment was collected through a museum exhibit where participants had their genotype taken and participated by trying different flavors and recording their results. A linear mixed effects model was used to model the results.

Colorado State University School of Veterinary Sciences: Effects of Social Media on Equine Health

- Consulting work for the School of Veterinary Sciences where researchers were investigating the impact of social media websites on equine health practices by analyzing the source of where horse owners were getting their equine health information. The data was analyzed using a multinomial model with random effects.

Data 2 Policy Project: Predicting the Proportion of People in Poverty by Denver Neighborhoods

- Fifty demographics from 2014 Denver Housing and Crime Incidents summed from Denver Police data were merged and analyzed using Ordinary Least Squares, Ridge, Lasso and Partial Least Squares machine learning methods. Goal was to see how well poverty levels could be predicted and which demographics contributed the most.

Clustering Credit Card Customers: “How to” Blog Post on K-Means Clustering

- Wrote a “how to” blog post on K-Means Unsupervised learning algorithm and coded an example in both R and Python

Sleep Deprivation Study: Linear Mixed Effects Model of Reaction Time

- Study was conducted to see the effect of 3 hours of sleep on 18 different subject’s reaction time. Data was analyzed using a linear mixed effects model with days without sleep as the fixed effect and subject as the random effect.

Bayesian Modeling of Diabetic Patients at Risk of Complications: Bayesian logistic regression model used to find the factors that contribute to diabetic patients with a high A1C

- Data set consisting of 394 diabetic patients was used to model diabetic patients at risk of complications where an A1C greater than 7 was considered at risk.
- Bayesian logistic regression model was used to determine which factors such as age, height, weight, cholesterol, blood pressure, etc., contributed to an A1C number greater than 7.

Predicting Miles Per Gallon: Predicting a car engines miles per gallon based on car specifications

- Project analyzed how accurately a car’s miles per gallon rating could be predicted using specifications such as cylinders, engine displacement, weight, etc.
- Supervised machine learning methods Lasso and Ridge Regression were used to analyze the data set.