Clément Mugenzi

Email: cmugenzi01@gmail.com
Cell Phone: (XXX) XXX
Portfolio: Website
LinkedIn: Profile

EDUCATION

Columbia University, New York, NY - ***Master of Science, Biostatistics*** 05/2021
***Relevant Coursework*:** Data Science, Statistical Methods, Statistical Inference, Causal Inference
University of Central Arkansas, Conway, AR **-** ***Bachelor of Science, Chemistry*** (Honors) 05/2016

SKILLS

Programming: Python, R, SAS, SQL, Hive, Hadoop, GCP (BigQuery & Vertex AI)
Technologies: Pandas, NumPy, Scikit-learn, Matplotlib, Tidyverse, TidyModels, Caret, R Markdown, R Shiny
Deep Learning: CNN, RNN (LSTM) with TensorFlow/PyTorch.
Machine Learning: Logistic Regression, Random Forest, XGBoost, SVM, Clustering, PCA, L1 & L2 Regularization
Statistics: A/B Testing (Experiment Design), GLM, Causal Inference, Longitudinal Data Analysis, Survival Analysis

INDEPENDENT RESEARCH PROJECTS

1. **Green Taxi: Predicting Percent Tip**

Machine learning tools in Python were used to predict percentage tip a driver would expect on each trip. Several features both categorical and continuous were considered as independent predictors such as total amount paid, trip distance, payment type, speed, etc. This project followed four main sections: Data Cleaning, Feature Engineering, Exploratory Data Analysis, and Model Building. A Gradient Boosting classification model was optimized (with a 96.1% AUC and 96.6% Accuracy) to predict whether or not a tip was provided, followed by a regression Random Forest model which then estimated the percentage tip given the tip was provided (MSE was 0.8).
See sample code and analysis here: Green Taxi Project

WORK EXPERIENCE

Aetna, a CVS Health Company, New York, NY 06/2021 – Present
Senior Data Scientist

- To make care management more efficient, I built a XGBoost algorithm to identify members with highest likelihood to pick up a care planning call from Aetna's nurses using members' past call history and recent clinical events.
- Optimized a LightGBM model to predict the likelihood of an inpatient event in the next 3 months for Dual Eligible special needs members and used Shapely values to reduce the dimensionality of their Health Risk Assessment.
- Conducted a thorough observational evaluation study for Aetna's Integrated Care Management program where I used Propensity score weighting for variable adjustment and Difference-In-Difference for effect estimation of the program.
- Using Rshiny, I helped build internal visualization tools that communicated program evaluation results and key performance metrics.

Columbia Heffner Biomedical Imaging Lab, New York, NY 06/2020 – 09/2020
Research Assistant

- Worked with CT images from COPD patients and Extracted, Transformed, and Loaded the dataset in R to run aggregates and plot different visualizations such as a Sankey Diagram and a bar chart.
- Conducted A/B Testing for contingency tables where I tested for independence between features and validated my results by conducting a permutation test.
- Built a Multinomial Log-linear model to study the association between emphysema subtypes from baseline to follow up and provided the model's interpretations.
- A regression model for the prediction of percent emphysema pixel was built in R by optimizing a Multivariate Adaptive Regression Spline (MARS) algorithm which resulted in a 90% score.
- Utilized Machine Learning tools in Python to again predict the percent emphysema pixel by optimizing XGBoost algorithm which resulted in a 82% score.

PUBLICATION

Synthesis and characterization of divalent metal complexes with bipyridylamide ligands, Clement Mugenzi *et al*, Journal of Coordination Chemistry 2015.

- Synthesized 2D and 3D polymers and used CrystalMaker to analyze their structure.
- Using CrystalMaker as repository, I surveyed the literature to compare our discoveries with previous researches in order to better understand our polymers' crystal structure.
- CrystalMaker is a visualization tool that builds, displays, and helps manipulate all kinds of crystal structures.
- **Transferable skills:** Computation, Critical Reasoning, Research Design, Data Visualization.
- **Reference:** Research Paper as First Author

PROFESSIONAL ORGANIZATIONS

- Member: Health Analytics Club - Columbia University 2019 – 2021
- Member: Computer Club – Biostatistics Department 2019 - 2021
- Member: American Chemical Society 2013 - 2016
- Member: Inspire Scholars Foundation 01/2014 - 07/2014

AWARDS & HONORS

- Taub Institute Award at Columbia 2019
- The Nicole Wable Hatfield scholarship 2014
- Rwanda Presidential Scholar (Provided to top 50 students from Rwanda) 2012-2016