Clément Mugenzi

Email: cm3928@cumc.columbia.edu
Cell Phone: (347) 703 2776
Portfolio: Website
LinkedIn: Profile

EDUCATION

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

SKILLS

Programming: Python, R, SAS, SQL, LaTeX
Technologies: Pandas, Numpy, Scikit-learn, Matplotlib, Tensorflow (keras), Tidyverse, Caret, R Markdown, R Shiny
Deep Learning: CNN, RNN (LSTM) with Tensorflow/PyTorch.
Machine Learning: Random Forest, XGBoost, SVM, Clustering, PCA, L1 & L2 Regularization
Statistics: A/B Testing, GLM, Time Series, Survival Analysis, Longitudinal Data Analysis, Logistic Regression

INDEPENDENT RESEARCH PROJECTS

1. **Emotions: Timeseries Analysis**
Machine learning tools in Python were used to predict emotional sentiments from brainwave readings. The Electroencephalography (EEG) dataset contains electronic brainwave signals from an EEG headset and was in temporal format. A classification model with an accuracy of 94% and loss (MSE) of 0.2, was finally built by optimizing a 1D convolutional neural network (CNN) algorithm.
See sample code and analysis here: Timeseries Project

2. **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

Columbia University Biostatistics Department, New York, NY 09/2020 – 12/2020
Graduate Teaching Assistant, Relational Database Systems
- Provided guidance to graduate students for course focusing on SQL coding and analysis, ensuring understanding of highly technical material for students from a range of educational backgrounds.
- Accurately provided grading and feedback for weekly assignments.
- Mentored students in building a relational database system as a final project.
- Collaborated with other teaching assistants to improve students' class experience.

Columbia Heffner Biomedical Imaging Lab, New York, NY 06/2020 – 09/2020
Statistical Analysis 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.

American Interplex Corp., Little Rock, AR 04/2017 – 11/2017

Data Analyst

- Built calibration curves (to construct analytical methods), using regression analysis, from which the instruments would infer and quantitate unknown samples.
- Devised a model to detect and/or drop outliers and together with a proper maintenance of the instrument contributed to reducing the processing time by 50%.

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 – Present
- Member: Computer Club – Biostatistics Department 2019 - Present
- Member: American Chemical Society 2013 - 2016
- Member: Inspire Scholars Foundation 01/2014 - 07/2014

AWARDS & HONORS

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