JOSEPH TRAN

tranjoseph11@gmail.com (484) 538-5604 Philadelphia, PA joseph-l-tran JLtran11

SUMMARY

Process oriented data analyst experienced in cleaning, interpreting, and analyzing data to support RCTs in a developing nation while using a diverse set of tools to furnish data visualizations, summaries, and reports that support policy recommendations.

SKILLS

PYTHON: Pandas, Scikit-lean, Numpy, Seaborn, Stats_models
MACHINE LEARNING: Classification, regression, clustering, feature engineering/Selection
DATA VISULAIZATION: Tableau, Matplotlib
SQL: Joining, funnel analysis, sub-queries, temporary tables
EXCEL: pivot tables, v-lookup, formulas/functions
STATISTICAL ANALYSIS: A/B testing, ANOVA, chi-squared, t-test, ARIMA

PROJECTS

Music Recommender Mar. 2020 - Aug. 2020

- Sourced #nowplaying-rs dataset of 17m listening events to design a music recommender system
- Data was cleaned, and three separate files representing listening event log, associated hashtags, and sentiment scores were combined into one data frame
- Matrix factorization algorithms from ALS and lightfm are used to find latent features of the user/item matrix to make user predictions
- Precision @ k scores are compared between models with lightfm and ALS model scoring 0.125 and 0.068, respectively
- Both models suffer from the 'cold start' problem and could be improved by creating a hybrid model with meta-data

Loan Default Predictor Sept. 2020 - Current

- Utilized Imperial College of London's loan default dataset of 150k samples and 755 features to develop loan prediction model
- Filter and ensemble feature selection methods are used in addition to dimensionality reduction to find an optimal set of features for the model
- Datasets from the results of the 3 methods are used for modeling
- Various classification algorithms are tested to find the optimal model and parameters via random_searchCV
- Models are benchmarked against an XGBoost classifier; Logistic Regression used along with ensemble methods.
- Best model was found to be XGBoost with SMOTE over-sampling yielding 0.677, 0.635 train and test ROC scores

EXPERIENCE

KH Global Health Consulting, *Data Science Fellow*, Apr. 2000 - Oct. 2020

- Provided data cleaning, management, and reporting services for health systems geographic data from Sierra Leone
- Led development of automated data file ingestion with python and glob; reduced load time from 1 week to several minutes
- Created and maintained Sankey diagrams to track importation and distribution of pharmaceutical products
- Ad hoc statistical analysis on health system data

Springboard, *Data Science Fellow*, Philadelphia, PA Feb. 2020 - Feb. 2021

Student in an online program consisting of 500+ hours of hands-on curriculum, with 1:1 industry expert mentor oversight, and completion of 2 in-depth capstone projects. Mastering skills in Python, SQL, data analysis, data visualization, hypothesis testing, and machine learning.

Freelance Consultant, *Project Manager* Jan. 2018 - Jan. 2019

- Coordinated with stakeholders to remodel and relaunch independent motel into a franchise destination
- Negotiated deadlines and budgets with local property developers and tradesmen
- Synchronized expectations of all involved parties with weekly budget updates and progress tracking with Gantt chart
- Project completed two weeks ahead of schedule, 6% under budget

Robson Forensic, *Mechanical Engineer Intern*, Lancaster, PA May 2016 - Aug. 2016

- Created finite element model to study stress and deflection as a result of wind interactions; hand calculations within 7 percent of computer model
- Responsible for design calculations and product development for intellectual property claims
- Field-guided inspections under expert oversight
- Produced mechanical drawings relating to intellectual property claims
- Assisted with evidence collection, documentation relating to forensic engineering field

Istar Productions, *Audio Engineer*Boston, MA / Las Vegas, NV July 2005 - Mar. 2010

- Sought out, and hired professional musicians for live client performances
- Nurtured client relationships to identify desired texture and sound within specific genres
- Served as client relationship associate to ensure final product fit within the artist's musical vision
- Invested in client outreach, acquisitions led to 43% revenue increase for 2008
- Musical Director for client live performances and multi-state tours
- Used Pro Tools HD and a wide array of VST plug-ins in conjunction with hardware outboard processing units to mix audio to the industry standard

EDUCATION

SUNY Maritime Sept. 2014 - May 2017
B.E Mechanical Engineering

Lehigh Carbon Community College Sept. 2010 - May 2013
A.S. Mechanical Engineering