CHIEMI KATO

Data Scientist

PROFILE:

In the past few years, I have built up my credibility on one simple principle: Work smarter, not harder. I have a hunger for learning new technologies and I strive to boost the productivity of everyone around.

TECHNICAL SKILLS

LIBRARIES:

- **Python**: NumPy, SciPy, Matplotlib, Pandas, Scikit-Learn, Keras, Tensorflow, NLTK, Seaborn
- **R**: Dplyr, MASS, Boot, Class, glmnet, Caret, randomForest, ROCR, ggplot

MODELS/METHODOLOGY:

- Linear Regression, Logistic Regression, KNN, Ridge, Lasso, SVM, Decision Trees, Random Forests, Neural Networks, Cluster Analysis

APPLICATIONS:

- Statistical Analysis, Web Scraping, Natural Language Processing, Time Series Analysis, Classification, Regression, Topic Modeling, Cross-Validation

LANGUAGES:

- Python
- R
- SQL
- SAS

GET IN TOUCH!

650-248-1342
568 2nd Avenue,
San Francisco, CA 94118

ckato@mail.sfsu.edu

WEBSITE/PORTFOLIO

chiemiko.github.io

EDUCATION

San Francisco State University
Class of 2019, GPA: 3.99

Bachelor of Science in Applied Mathematics | Graduating Winter 2019

- Related coursework: Categorical Data Analysis, Statistical Learning & Data Mining, SAS Programming, Advanced Python, Numerical Analysis, Data Structures and Algorithms, Probability and Statistics, Mathematical Modeling

Hackbright Academy

Python Programming Bootcamp | Fall 2018

- Intensive coursework in Python programming in topics including control flow, object- oriented programming, comprehensions and data structures
- Constructed computerized Alarm Clock as part of final course project

EXPERIENCE

Data Science Intern

MARi | Jul 2019 - Present

- Responsibilities include implementation of topic modeling and natural language techniques to job posting data in order to build automated skill mappings for professional development

Machine Learning Research Assistant

Tao He's Laboratory (in collaboration with Fong Lab, UCSF) | Jun 2019 - Present

- Developing a predictive classification system in order to identify existence of cancer cells in T-Cell Receptor sequences to measure effects of immunotherapy
- Use of deep learning (CNN, RBF, and MLP) and supervised learning algorithms (SVM, Logistic Regression, Boosting, Bagging, Random Forests, Lasso Regression) to propose a comprehensive model for deployment in future medical innovation research

Research Assistant for Federico Ardila in Algebraic Combinatorics Project

SF State Department of Mathematics | Dec 2018 - May 2019

- Evaluated the eigenvalues and eigenvectors of tropical Laplacians that arise from four families of tropical surfaces of root polytopes to confirm the generalized Hodge Conjecture
- Wrote Python programs and built algorithms to generate data for analyzing patterns
- Proved theorems and presented findings to other researchers in the community

ACADEMIC PROJECTS

Instacart Market Basket Analysis
Kaggle Project

SF State Statistical Learning & Data Mining Project | Spring 2019

- Processed high quantities of Instacart historical grocery data to build predictive classification models based on the likelihood of users reordering certain products in the future
- Extensive experience with supervised learning methods (Logistic Regression, Decision Trees, SVM) to identify models with highest classification accuracy

AWARDS

- Bengier Foundation University Scholarship, SFSU Office of Financial Aid, May 2019
- SFSU Dean's Honors List, December 2018 and June 2019
- David Meredith & Friends Scholarship, SFSU College of Science & Engineering, *October 2018*