

LinkedIn: [linkedin.com/in/jiaming-chen-data-analyst/](https://www.linkedin.com/in/jiaming-chen-data-analyst/)

GitHub: github.com/jmistr1

Personal Page: jmistr1.github.io

JIAMING CHEN

San Diego, CA 92126 • (928) 788-5045 • jchen1105@hotmail.com

SKILLS

Programming: Python(Keras, Pandas, sklearn, NLTK), R(glmnet, caret, ggplot2, tree), C, C++, MATLAB, SQL

Models: Smoothing Splines, Logistic Regression, A/B testing, Bagged Tree, Boosted Tree, SVM, CNN, NLP, K-Means, Collaborative Filtering Recommender, Matrix Factorization, Particle Swarm Optimization, Word2Vec, Autoencoder, Wavelet Decomposition

Tools: PySpark(SparkSQL, MLlib), AWS, OpenCV, Linux, Latex, Excel

Languages: French, English, Mandarin Chinese

EXPERIENCE

OPERA SOLUTIONS

San Diego, CA

Senior Analyst Specialist

08/2018 – now

- Extract healthcare data from MySQL server, conducted data ETL and data profiling on AWS server using PySpark
- Performed data visualization of healthcare data to build baselines of member and provider behavior outlier detection
- Used Autoencoder model to reduce the dimensions of provider medical claim data. Built outlier detection model using HDBSCAN to flag healthcare providers with abnormal behavior.
- Trained Word2Vec, Doc2Vec (gensim) model on 9 GB member claim sequence data. Deployed “Sequence to Sequence” model to reduce the dimension and vectorize members’ medical record data.
- Trained Convolutional Autoencoder, Variational Autoencoder and neural image caption generator model to vectorize medical claim record data. Compared the performance of different model using t-SNE and k-Nearest Neighbor.
- Assembled provider outlier detection and member clustering algorithm to a scoring system for a medical insurance company.

ECALCHARGE INC

Berkeley, CA

Data Scientist Intern

06/2017 – 08/2017

- Parsed and extracted electrical vehicle driving history raw data from data provided by car manufacturer on AWS
- Cleansed missing values and outliers and used regression to impute the missing driving event in the data using Python
- Performed data visualization of driving patterns to find informational features
- Performed feature engineering and feature selection, used PCA to reduce the dimensionality and reduce
- Applied k-means clustering to identify similar driving patterns, used elbow method to find optimal number of clusters.

EDUCATION

NORTHERN ARIZONA UNIVERSITY

Flagstaff, AZ

Master of Science in Electrical Engineering, Sept 2016 - July 2018

GPA: 3.9/4.0

- Pattern Recognition, Large Scale Data Structures, Statistical Computing, Image Processing
- Academic research in biomedical signal processing and automated classification system design using kernel method and multiobjective optimization method

SOUTHWEST JIAOTONG UNIVERSITY (SWJTU)

Chengdu, China

Master of Science in Electrical Engineering, Sept 2015 - Jan 2018

GPA: 3.4/4.0

ECOLE CENTRALE MARSEILLE (double degrees program with SWJTU)

Marseilles, France

Diplôme d'Ingénieur (equivalent to Master Degree), Sept 2013 – May 2015

GPA:3.4/4.0

SOUTHWEST JIAOTONG UNIVERSITY (SWJTU)

Bachelor of Electrical Engineering, Sept 2011 - June 2015

PROJECT

ADVANCED WARNING AND CLASSIFICATION SYSTEM FOR ECG SIGNAL

08/2016 - 08/2018

- Designed ECG signal acquisition system with body sensor and wireless transmission module
- Implemented ECG signal preprocessing and segmentation algorithm using wavelet decomposition. Extracted features from ECG signal and performed feature selection
- Designed a two-staged hierarchical classification system including weighted k-Nearest Neighbor and a designed clustering algorithm for advanced warning of cardiovascular disease
- Improved the system performance with a designed feature space reshaping algorithm and Multi-Objective Particle Swarm Optimization.

LinkedIn: [linkedin.com/in/jiaming-chen-data-analyst/](https://www.linkedin.com/in/jiaming-chen-data-analyst/)

GitHub: github.com/jmistrall

Personal Page: jmistrall.github.io

MUSIC PLAYER CHURN PREDICTION

02/2018 - 03/2018

- Used Shell Script and Spark parse user behavior log file(16GB), cleaned log file, resampled unbalanced classes, visualized data and performed Exploratory Data Analysis using PySpark and Pandas
- Performed Feature Engineering (including frequency, recency, acceleration, etc.), generated user churn label and performed classification using random forest and multi-layer perceptron achieving AUC of 0.89.

YELP DATASET CHALLENGE: NLP & SENTIMENT ANALYSIS

01/2018 – 02/2018

- Cleansed and filtered review metadata(3.28GB) and joined it with relevant tables using Python(Pandas)
- Used NLP techniques (e.g. TFIDF Vectorizer) to extract features from unstructured review text data with sklearn, nltk
- Clustered the reviews with KMeans clustering and fitted Naive Bayes Classifier to predict user sentiment
- Built a restaurant recommender system using Collaborative Filtering and Matrix Factorization based on user's ratings

LENDING CLUB RISK ADJUSTED INTEREST RATE PREDICTION

12/2017 – 01/2018

- Extracted 12 features from raw lending club loan data containing different types, such as categorical, numerical and time series data, imputed missing data using multivariate imputation by chained equation (MICE) algorithm
- Performed feature selection through exploratory analysis
- Fitted linear regression model with regularization to control for multicollinearity and also built decision tree, random forest, boosting decision tree to predict interest rate for each loan
- Achieved 0.3 RMSE by boosting decision tree model on test data set.

KAGGLE ICEBERG IMAGE CLASSIFICATION CHALLENGE

10/2017 – 12/2017

- Performed image denoising and image augmentation on the SAR images using OpenCV and Python(Keras)
- Applied Convolutional Neural Network on the training images and tested performance using Keras

PUBLICATIONS

-
1. J. Chen and A. Razi, "A predictive framework for ECG signal processing using controlled nonlinear transformation," *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, 2018, pp. 161-165.
 2. J. Chen and A. Razi, "Remote ECG Monitoring Kit to Predict Patient-Specific Heart Abnormalities", *Journal on Systemics, Cybernetics and Informatics (JSCI)* Volume 15 - Number 4 - Year 2017, pp. 82-89