LinkedIn: linkedin.com/in/jiaming-chen-data-analyst/
GitHub: github.com/jmistral
Personal Page: jmistral.github.io

JIAMING CHEN

San Diego, CA 92104 • (928) 788-5045 • jchen1105@hotmail.com

SKILLS

Programming: Python(Tensorflow, PyTorch), SQL, Shell Script, R, C++, MATLAB,
Models: Logistic Regression, A/B testing, Boosted Tree, SVM, CNN, RNN, NLP, Transformer, K-Means, t-SNE, DBSCAN, Spectral Clustering, Collaborative Filtering Recommender, Matrix Factorization, Word2Vec, Autoencoder
Tools: PySpark, Flask, Qt Builder, AWS (SageMaker, RedShift), OpenCV, Linux, Latex, Confluence, JIRA
Languages: French, English, Mandarin Chinese

EXPERIENCE

AV LABS San Diego, CA
Algorithm Scientist 02/2019 – Now

Natural Language Processing Research Project:
- Performed data cleaning and pre-processing to extract training and test dataset including 10M corpus and 2000 pairs of multiple choice medical questions from medical literatures
- Pre-trained, fine-tuned models (BERT, XLNet) to answer medical questions, improved accuracy from 20% to 41%
- Designed a web interface held on local work station for internal model testing using Flask

Recommendation System
- Extracted and processed consumer transaction data from client's database using SQL
- Performed feature extraction, dimension reduction on transaction data, created Eigen Customer visualization
- Used HDBSCAN to cluster customers and built a recommendation system based on clustering result

GIS and UAV System
- Designed and developed frontend of GIS software plugin for infrastructure inspection
- Developed an algorithm for automated UAV flight path planning, and image alignment algorithm using OpenCV

Medical Device Consumer Data Analysis
- Combined medical insurance claim data with client's device usage data on RedShift, conducted feature extraction and patients clustering based on clinicians' feedback and client's requirement.
- Performed data visualization on patient clusters, defined target variable to measure benefit from device usage
- Build predictors for benefit and delivered explainable analysis report to non-technical leadership team
- Processed and imputed device usage data, extracted time series features and built a churn prediction model for client

ELECTRIFAI (FKA OPERA SOLUTIONS) San Diego, CA
Senior Analytics Specialist 08/2018 – 02/2019
- Extracted healthcare data from MySQL server, conducted data ETL and data profiling using PySpark and shell script;
- 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 to reduce the dimension and vectorize patients' medical record data
- Clustered patients using agglomerative clustering, validate the results using clinicians provided test dataset
- Assembled provider outlier detection and member clustering algorithm to a scoring system for the client

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;
- Removed outliers and used interpolation to impute the missing driving event in the data using Python;
- Performed feature engineering and feature selection, used PCA to reduce the dimensionality;
- 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 - August 2018 GPA: 3.9/4.0
- Pattern Recognition, Large Scale Data Structures, Statistical Computing, Image Processing

ECOLE CENTRALE MARSEILLE (double degrees program with SWJTU) Marseilles, France
Diplôme d'Ingénieur (equivalent to Master Degree), Sept 2013 – May 2015 GPA:3.64/4.0
SOUTHWEST JIAOTONG UNIVERSITY (SWJTU) Chengdu, China
Bachelor of Electrical Engineering, Sept 2011 - June 2015 GPA: 3.61/4.0