Pengchong Tang

Kalamazoo MI 49009 **Email:** ferrarisf50@gmail.com **Phone:** (269) 365-6057 Github Linkedin

SKILLS

Programming languages:	SAS, R, Python, SQL, MATLAB, Scala, JavaScript, C++, C#, VBA
Web Technology:	HTML, XML, JSON
Data science & big data tools:	Spark, MLlib, PySpark, Tableau, Sklearn, H2O.ai, AWS, SPSS
Version control systems:	Git, SourceTree
Databases:	Oracle, SQL Server 2012
Technology knowledge:	Machine learning, Data Analysis, Data Mining, Data Wrangling, Predictive Modeling, Statistics, Natural Language Processing (NLP), Operations Research, Experimental Design

PROJECTS

Credit Card Fraud Detection — 11/2017

- Explored and analyzed a credit card transaction dataset using R ggplot2 and built a classification model using Sklearn in Python. The model achieves 0.8 score on average on AUPRC, which is able to detect about 80% of frauds.

Udacity Tweeter Data Wrangling — 12/2017

- Gathered 6000 tweets of @dog_rates through Twitter API, assessed the data quality and programmatically cleaned data to ensure completeness and tidiness.

NYC Taxi Trip Duration — 01/2018 - present

- Gathered 1.4 million NYC taxi trip routes information during the first half of 2016 from OSRM, assessed and cleaned the data with specific criteria, visualized and analyzed the data using Tableau, implemented machine learning algorithms (Xgboost) to build a model to predict taxi trip duration, scoring 0.445 on the private leaderboard.

EXPERIENCE

Eurofins Lancaster Laboratories / SAS programmer at Kalamazoo, MI **11/2014 – present**

- Support clients with clinical data management, define validation rules according to the data management plan (DMP), design machine learning models and portable Python apps to automatically collect information, generate SAS templates, process 30+ lab data requests per week, including data import, data correction and data flagging, saving 60% of processing time. (Main skills: SAS/MACRO, SAS/BASE, SAS/SQL, Sklearn, NLP, Random forest, SQL Server)
- Revamp existing programs and develop new apps to automatically retrieve data through the Electronic Data Capture (EDC) System API, map data to the CDMS database daily for 20+ clinical studies, reducing labor hours for clients by 90%. (Main skills: XML, JSON, Python, SAS)
- Develop and design tools and apps to automate the repeated data management tasks, including email management and form filling, successfully reducing labor hours for clients by 80%. (Main skills: VBA, Python, Web scraping)

EDUCATION

- MS in Statistics, George Washington University, Washington DC **09/2010 – 06/2012**
- BS in Mathematics, Sun Yat-sen University, Guangzhou, China **08/2006 – 06/2010**
- Udacity Data Analyst Nanodegree Program **11/2017 – 01/2018**

CERTIFICATIONS

- SAS Certified Advanced Programmer for SAS 9 **03/2014**
- SAS Certified Base Programmer for SAS 9 **11/2012**
- R Programming (Coursera) **03/2015**
- Data Visualization and Communication with Tableau (Coursera) **11/2017**

LANGUAGES

Chinese Mandarin(native), Cantonese(native)