

# Pengchong Tang

Email: [ferrarisf50@gmail.com](mailto:ferrarisf50@gmail.com) Phone: (269) 365-6057 [Github](#) [Linkedin](#)  
Kalamazoo MI 49009

## SKILLS

Programming languages: SAS, R, Python, JavaScript, C/C++, C#, SQL, MATLAB, Mathematica, HTML  
Software and tools: MS Excel, Oracle, Tableau, SQL Server, Git, H2O.ai, Spark  
Technologies: Machine learning, Data Analysis, Data Mining, Data Wrangling, Predictive Modeling, Statistics, Natural Language Processing (NLP), Operations Research

## EXPERIENCE

**Eurofins Lancaster Laboratories / SAS programmer** 11/2014 - present

- Work as a coordinator with data manager, biometrician and study representative on supporting multiple Veterinary Medicine Research and Development (VMRD) projects.
- Work with study personnel to create Data Management Plans (DMP), provide study-specific setup, modify SAS template to define validation checks, perform data import, data correction, data transformation and data extraction according to business standards. (Main skills: SAS/MACRO, SAS/BASE, SAS/SQL)
- Support clients in maintenance of Clinical Data Management System (CDMS). Create SQL script to query and update the SQL database. (Main skills: SQL Server, XML)
- Develop a smart Excel reader for spreadsheet processing using Python and NLP techniques, which is able to automatically capture variable names and data on lab sheets. Design Python programs to create SAS scripts and automate CDMS daily maintenance tasks. (Main skills: scikit-learn Pandas NumPy)
- Develop and implement SAS programs and web scraping scripts to perform data mapping from Electronic Data Capture (EDC) System to the CDMS database.

## PROJECTS

**Kaggle.com Credit Card Fraud Detection**

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

**Udacity Tweeter Data Wrangling**

- Gathered thousands of @dog\_rates tweets through Twitter API, assessed the data quality and programmatically cleaned data to ensure completeness and tidiness.

**Kaggle.com NYC Taxi Trip Duration**

- Gathered 1.4 million routes information from OSRM, assessed and cleaned data with specific criteria.
- Designed a Tableau data visualization story about factors that affect the taxi trip duration during Jan2016 to Jun2016.

## CERTIFICATIONS

|   |         |
|---|---------|
| SAS Certified Advanced Programmer for SAS 9 (Serial Number: AP010757v9) | 03/2014 |
| SAS Certified Base Programmer for SAS 9 (Serial Number: BP028383v9)     | 11/2012 |
| R Programming (Coursera)  | 03/2015 |
| Data Visualization and Communication with Tableau (Coursera)            | 11/2017 |

## EDUCATION

|   |                 |
|---|-----------------|
| Udacity Data Analyst Nanodegree                               | 10/2017-01/2018 |
| MS in Statistics, George Washington University, Washington DC | 09/2010-05/2012 |
| BS in Mathematics, Sun Yat-sen University, Guangzhou, China   | 09/2006-06/2010 |

## LANGUAGES

Chinese Mandarin(native), Cantonese(native)