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, PySpark, Tableau, Sklearn, H2O.ai, AWS, SPSS
Version control systems:	Git, SourceTree
Databases:	Oracle, SQL Server 2012
Knowledge:	Machine learning, Data Analysis, Data Mining, Data Wrangling, Predictive Modeling, Statistics, Natural Language Processing (NLP), Operations Research

PROJECTS

Kaggle.com 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 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   01/2018 - present
• 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 Jan. 2016 – June 2016.

EXPERIENCE

Eurofins Lancaster Laboratories / SAS programmer at Kalamazoo, MI   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.

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)

Mathematica, HTML

Software and tools: MS Excel, Oracle, Tableau, SQL Server, Git, H2O.ai, Spark
Technologies: