Yunkun Yang

Contact Information : harris.yang@uwaterloo.ca | 519 - 781 - 9052

LinkedIn: linkedin.com/in/yunkunyang611 | Link
GitHub: github.com/harrisyang611/Harris-Yang-611 | Link

SUMMARY OF QUALIFICATIONS

- Data Science, Statistics and Programming: R, Python, TensorFlow, Scikit Learn.
- Database & BI tools: SQL, NoSQL SQL Server, BigQuery, Redshift, MongoDB.
- Visualizations & BI tools: Tableau, Power BI, ggplot, matplotlib, seaborn.
- Cloud Environment: Amazon Web Service(AWS), Google Cloud Platform(GCP), Watson Studio, Boto3, google.cloud.
- Five years of experience as data analyst and data scientist including three years under professional settings.

PROFESSIONAL EXPERIENCES

Machine Learning Engineer & Researcher, Shop Bonsai   **May – Aug 2019**
- Researched Minimum Viable Product for image & video classification and derived business decision.
- Reproduced image classification algorithms from multiple research papers on own initiative.
- Used GCP to train machine learning models and provided consequential analysis and visualization.

Data Analyst, Product & Analytics, Rubikloud   **Sep – Dec 2018**
- Built automated process for data ingestion and designed dashboards for different clients with Tableau and Redshift.
- Communicated with product team to design better data quality testing.
- Conducted ad-hoc data extraction using PySpark and AWS.

Data Analyst, Internal Audit Data Analytics, Scotiabank   **Jan – May 2018**
- Initiated fuzzy string matching algorithm for the team to generate general mapping tables depending on different string similarity metrics.
- Web scraped public data through Python for Trade Finance AML project.
- Implemented automated ETL pipeline to migrate data between Lotus Note and SQL Server using R and SQL.

Data Analyst, Internal Audit Data Analytics, Scotiabank   **May – Aug 2017**
- Used decision tree and random forest to analyze Security Risk Rating for Wealth Management team.
- Efficiently communicated with auditors to help them understand applications of machine learning.
- Detected and fixed data issues caused by system updates.

PROJECTS

Battle of Neighborhood, GitHub Repo   **Aug 2020**
- Scraped data from Wikipedia, analyzed, modelled geospatial data using Python and Machine Leaning algorithm in Watson Studio and wrote report and published article in Medium for final presentation.

Microsoft Malware Prediction   **Apr 2019**
- Applied different machine learning techniques and used LGBM as final model after analyzing various performance indexed from models.

HK Retailor Dashboard Projection   **Oct 2019**
- Built a KPI dashboard with automated data ingestion and processing procedure using Python and AWS for the biggest HK health and beauty retailor along with created the QA process for the dashboard and used Tableau to automate the metric calculation and validation.

Data Quality Report Shiny App, GitHub Repo   **Apr 2019**
- Optimized the program flow and upgraded the test plan with different features then designed a self-served drag-and-drop R Shiny App, producing DQ reports for auditors.

EDUCATION

Bachelor in Statistics, University of Waterloo   **2015-2020**
- Courses taken: Computational Analysis, Experimental Design, GLM, Computational Inference, Statistical Learning, Data Visualization, Forecasting (Time Series Analysis), Function Estimation

Data Science Professional Certificate, IBM   **2020**

ACTIVITIES AND INTERESTS

Lyrics Composer and Rapper   2013 – Present
- Published over 35 raps & remix