Runye (Gabriel) Wang

560 West 43rd St ,10036 Phone: 706-362-3561   E-Mail: me@gabrielwry.com

Website: http://gabrielwry.com/

Data Scientist with 3+ years of project and professional experience. Delivers **end-to-end Big Data solution**. Skilled in **Data Visualization, Data Mining**, **Data Engineering** and **Machine Learning.**

Skills

Programming Languages:	Data Science:	Dev-Op:
Professional - Python, SQL;	Jupyter, Data Visualization, Data Mining;	Django, RESTful API, Google Cloud Platform;
Project - Java, C, MatLab, R	Machine Learning, Statistical Modeling	UNIX, Docker, Kubernetes

Professional Experiences

Associate Data Scientist at HEED LLC, New York, NY   2018.09 - 2019.07
Designed, implemented and delivered big data solution for creating content and feature in an AI driven sport analytics platform.

- Designed and prototyped **Data Warehouse** system for large sensor data set on **Google BigQuery** with scalable ETL process micro-services and worked with engineering team to deploy on **Google Cloud Composer**;
- Designed and implemented **Business Intelligent tools** using **Google Data Studio**, **Elastic Search Kibana** to support creative content team with real time data analysis during sport event;
- Analyzed app user activities with advanced **Data Mining** techniques like **decision tree learning** and **random forest** to support marketing team strategy decision and achieved 400% user growth from v1.0 to v2.0;
- Researched and implemented **rule-based system** that uses **Neural Network** trained sensor analytic data and social media metrics obtained from **Topic Modeling** and **Sentiment Analysis** to annotate user-interested sport moments.

Research Assistant at Emory University, Atlanta, GA   2017.10 to 2018.09
Conducted text analysis research and assisted software development, supervised by Prof. Roberto Franzosi.

- Implemented end-to-end **NLP** pipeline that assisted over 200 sociology students and researchers with quantitative text analysis tasks like **Word2Vec, Data Visualization** with **Gephi and Google Map**;
- Designed and implemented **anomaly detection** algorithms that pre-processed millions of Italian marriage records scanned from local newspapers for Sociology researchers to create papers based on it;
- Created and presented **regression model** to predict interviewee responses on over 130K interviews from UK personality development project, achieved 20% more accuracy than traditional sociology method;

Data Engineer Intern at CDC, Atlanta, GA   2017.06 to 2017.08
Actively involved in Agile development and Business Intelligence of phase one launch of public health data analytic product.

- Implemented and monitored automated ETL process on **AWS Redshift** to ingest over 100 million records of Southeastern US 2017 public health data;
- Worked with engineering team to integrate analytic tool powered by **pySpark** into **Django** application to assist the Data Analysis team with **time series** and **classification** analysis;
- Worked with researchers to validate data analytic result and purposed public health regulation;

Projects

Prediction on European Football Player

- A **Machine Learning system** to predict which European Football players are most likely to be successful in 5 years and achieved 84% accuracy, beat 95% participants;
- Joined the multiple data source to annotate over **20k** players with **50** attributes; Identify **Frequent Items Mining** and **Information Gain** to identify key attribute for players with top 10% market value from different **clusters**; apply **Time Series Forecasting** to predict young prospects' future market values;

Education

Emory University, Class of 2018, B.S in Computer Science and Engineering Physics