Runye (Gabriel) Wang

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

GitHub Link: https://github.com/gabrielwry

Junior **Data Scientist** with 3+ years of project and professional experience. Deliver **end-to-end big data solution** for company and achieved 400% user growth in an AI driven sport analytic platform. Skilled in **Data Mining**, **Data Engineering** and **Machine Learning.**

Skills	Programming Language:	Data Science:	DevOp:
	Python, Java (Professional)	SQL, BigQuery, Data Mining,	Restful, Django (Professional)
	C, R, MATLAB (Project)	Machine Learning, Statistical	JavaScript, Docker, Kubernetes MySQL,
		Modeling (Professional)	Elasticsearch, Linux (Project)

Professional Experiences

Associate Data Scientist at HEED LLC, New York, NY   2018.09 - Present
Design, implement and deliver high quality big data solution to content and feature creation in an AI driven sport analytics platform.
- Design and implement **Data Warehouse** system on **Google BigQuery** with scalable ETL process microservices and work with developer to deploy on **Google Cloud Composer** to ingest large dataset;
- Design and implement **Business Intelligent tools** using **Google Data Studio**, **Elastic Search Kibana** to support real time data analysis during sport event and content delivery to over 40K user world-wide;
- Analyze large sport dataset in MMA, Basketball and Soccer with advanced **Data Mining** techniques like **decision tree learning** and **random forest** to strategize innovative content creation that attracted 400% more users in two months;
- Research and implement **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 perform quantitative text analysis from **Word2Vec** to **Data Visualization**;
- 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 development 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 and research data analytic result and purposed public health regulation;

Projects

Kaggle 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