

# WEIXING LI

669-273-6878 | wli10@email.wm.edu | <https://www.linkedin.com/in/w-li/> | <https://github.com/jeremite>

◆ Pleasanton, CA 94566 ◆

## Experienced Data Scientist & Machine Learning Engineer

Versatile, results-driven professional with 2+ years of experience in big data engineering, machine learning pipeline building, and web app development. AWS machine learning certification. Skilled at solving business problems with large scale data using python/R, spark SQL. Proven problem solving and analytical thinking skills, a fast learner, and the ability to adapt to evolving industry trends.

### EXPERIENCE

#### Data Scientist – Verisk Analytics corporate, San Francisco CA

2019.8 – Present

- Built machine learning pipeline and Xgboost classifier for Insurance claim fraud detection on a workflow tool.
- Developed Spark SQL & Pyspark code to deal with around 20 TeraByte size data, running on AWS EMR.
- Reduced time required to generate features by 80% through building some Flask Web Applications.
- Published the Pyspark version Featuretools - Featuretools4spark to render feature engineering with Spark.
- Improved the metric by 10% by adding features from unsupervised algorithms like Autoencoder and EM clustering.

#### Ferguson Enterprise (Capstone Experience), Virginia

2019.3 – 2019.5

- Proposed a Multi-Channel Attribution Model to assess the impact of various advertising channels on conversion.
- Employed Attention + LSTM RNN and Embedding in Keras and Tensorflow on consumer behavioral and demographic data and boosted the evaluation performance to 0.9 AUC and 0.96 accuracy.

#### Analyst, Communication Strategy manager – Blue Focus International Group, Beijing

2016.11 – 2017.8

- Headed a project that attained 700,000 followers and decreased the average cost of acquiring fans by 80% through applying a new advertising strategy for China Mobile sub-brand - MIGU Facebook Page.
- Developed a classification Tree model with Python, raising the user engagement rate by 120%.
- Conducted A/B testing on the variants like different posting times and hashtags, etc., supporting the new strategy.
- Visualized and Illustrated the analysis effectively to stakeholders to help decide the budget allocation.

### PROJECTS

#### Automated Feature Engineering Web Application

2020.10 – Present

- Designed an interactive Feature Engineering web app to effectively save 90% data engineering coding time.
- Developed entire frontend and backend modules using Python and Flask Web Framework.
- Utilized the JQuery, Ajax, and NoSQL MongoDB database and bootstrap framework for responsive web pages.
- Made the app portable by using Docker-Compose to define and run multiple Dockers and also python installable.

#### Quora Insincere Questions Classification

2018.9 – 2018.11

- Constructed an NLP text anomaly detection model to detect insincere questions from Quora.com.
- Transformed 1 million + text data through Spark Pipeline using Scala on Databricks.
- Implemented merged Naïve-Bayes Logistic Regression, reducing training time by 90% compared to RNN.
- Authored an article distributed by curator on Medium.com, which illustrated the text mining and training;

#### AI-Student Attendance check

2018.10-2018.12

- Built a computer vision project for checking the attendance of students using word triggering and face recognition.
- Utilized GRU-based model to realize word triggering function and OpenCV to control the camera, collect face photos and realize face detection in the streaming video; Applied FaceNet to recognize identity with faces;

### CERTIFICATIONS

◆ AWS certified specialty – Machine learning ◆ Google Tensorflow Developer Certificate ◆ Linux Command Line certification; Big data specialization certification ◆ UCSC Silicon Valley extension: Data analytics and Database Management

### CORE SKILLS

◆ Python ◆ R ◆ Spark SQL ◆ SQL ◆ Scala ◆ Git ◆ Pandas ◆ Numpy ◆ Scikit-learn ◆ Flask ◆ NLP Text Mining and Computer Vision  
◆ NLTK ◆ Spacy ◆ Data Visualization ◆ Matplotlib ◆ Tableau ◆ Neural Network ◆ TensorFlow ◆ Keras ◆ Big Data ◆ Spark ◆ Hadoop  
◆ Deployment ◆ Flask ◆ Aws ◆ Sagemaker ◆ Time-series analysis, ◆ Pattern recognition ◆ Clustering ◆ Tree-based modeling  
◆ Linux Command Line

### EDUCATION

Master of Science, Business Analytics **William & Mary, Raymond A. Mason School of Business (GPA 3.95/4.00)**

Master of Science, Marketing specialized in Customer Management **University of Strathclyde**

