Senior Data Scientist Consultant Senior Data Scientist Consultant Senior Data Scientist - Consultant - Verizon Wireless Irving, TX Over 7 years hand on experienced in the areas of machine learning, deep learning and statistical modeling, big data analytics. Including prediction modeling, Time series forecasting, Financial Fraud Detection, OpenCV /Image Processing, Object Detection, Semantic Segmentation, Deep Speech Recognition, and NLP; Expertise in multiple deep learning platforms (Tensorflow, Keras, Intel Neon/BigDL, etc.) with diversified use cases; Experienced in Spark, Relational Databases, Graph Databases, NoSQL Databases, and related ML use cases. Authorized to work in the US for any employer Work Experience Senior Data Scientist Consultant Verizon Wireless - Irving, TX August 2018 to Present Data Science Projects for Wise of Customer Team at Verizon Wireless. Enhance online-sale performance and customer experiences based on text mining, recommendation systems and machine learning modeling. NLP and Text Mining for VOC(voice of customers-online chats): Topic Extraction with n-grams by LDA, NMF, LSI/SVD, and LDA2VEC; Dashboard visualization using pyLDAvis, t-sne, D3/js, etc. Sentiment and semantic analysis. Language modeling using spacy, gensim/word2vec, etc. Improve customer experiences and online purchase/visiting rates through integrations of recommendation system, prediction modeling, hypothesis tests, and A/B tests, etc. Perform under-sampling/super-sampling for unbalanced datasets. Fine tune the hyperparameters for major algorithms and evaluate the models.

Data science automated workflow on AWS/Docker, build Flask application to AWS Elastic Beanstalk, setup connection to the databases, build web interface to indicate desired database table and filtering keywords to extract data automatically, and to input modeling parameters in order to train the model on click, then return the modeling results/visualizations as dashboard to the webpage directly. Senior Data Scientist - Consultant Dell EMC - Round Rock, TX November 2017 to Present Artificial Intelligence Deep Learning Project: research on deep learning platforms with variant use cases, evaluating different deep learning platforms, infrastructure with Spark and GPU parallel systems. Testing wide variety of use cases: Lung Nodule Classification, Object Detection by SSD and YOLO, financial time series forecasting, Deep Speech Recognition, etc.. Testing Spark bigDL/Intel NEON/Tensorflow/CNTK/Keras, performance evaluations on backends of

CPU/MKL/GPU. Focusing on tensorflow and BigDL on spark/Hadoop. Presented the project use cases results to both external customers and Dell internal stake holders. Big data analytics with Hadoop, HiveQL, Spark RDD, and Spark SQL. Intel Deep Learning training and NVIDIA deep learning training. Senior Data Scientist Consultant Optum/United Health Group - Hopkins, MN March 2018 to August 2018 Data Science Innovative Project for United Health Group Big Data Apps Team: research on machine learning and deep learning algorithms to optimize business performance of health care/insurance. Apply machine learning and deep learning algorithms to match members/claims, Using Jellyfish, Difflib, Pandas, PySpark, and Scikit_Learn; NLP, SimHash, Fuzzy Matching Algorithms; Optimized the cross-table computations by Pandas functions and Spark. Performed clustering analysis to recognize patterns and identify outliers and anomaly for provider medical claims, using DBScan, Level Set Tree, Hidden Markov Model, and Self-Organizing Maps. Time series analysis to forecast number of health claims and total claim amount per period. Using Keras, Tensorflow, and ARIMA modeling for time series forecasting; Machine learning, Decision Tree and Recommendation Modeling on Graph Networks. Data transformation using Hive Spark SQL and Pandas, feature engineering based on Pandas and Spark RDD. Python Developer/Data Scientist Zero One Consulting - Katy, TX July 2015 to October 2017 Machine Learning Projects based on Python, R, SQL, Spark MLlib and SAS. Performed data exploratory, data visualizations, feature selections, and model validations. Improved the model performance on test data, for example: In a customer segmentation/campaign and credit risk prediction project, the metric scoring was lifted to 0.83 for final model from 0.57 for legacy model. Applications of machine learning algorithms, including kNN, NB, ANN, random forest and boosted tree, SVM, SGD (stochastic gradient descendent regression), PCA, Singular Value Decomposition, NLTK/textBlob, neural network, spark MLlib and deep learning using Tensorflow/Keras, and CNTK. Performed time series forecasting in R and Python Keras, natural language processing, statistical analysis, generated reports, listings and graphs. Big data analytics with Hadoop, HiveQL, Spark RDD, and Spark SQL. Tested python/SAS on AWS cloud service and CNTK modeling on MS-Azure cloud service. Geophysical Data Scientist Sclumberger Geosolution - Houston, TX December 2012 to

December 2016 Built prediction models of major subsurface properties for underground image, geologic interpretation and drilling decisions. Utilized advanced methods of big data analytics, machine learning, artificial intelligence, wave equation modeling, and statistical analysis. Provided exclusive summary on oil/gas seismic data and well profiles, conduct predictive analyses and data mining to support interpretation and operations. Cross-correlation based data analysis method through Python and R on multi-offset-well to help predict the models and pore-pressure ahead a little for real time drilling. Big data modeling with incorporation of seismic, rock physics, statistical analysis, well logs and geological information into the 'beyond image'. Using Python and Java, developing, operationalizing, and productionizing machine learning models to make significant impact on the geological pattern identification and subsurface model prediction. Analyzing seismic and log data with sub-group analysis (classification-clustering, hybrid approach by PCA and SOM) and model prediction methods (regression, random forest, boosted tree, standard neural network etc.). Use SAS statistical regression method to simulate the anisotropic trend. Tested the migrated data processing system on Google Cloud with velocity model updating tasks. ETL to convert unstructured data to structured data and import the data to Hadoop HDFS. Utilized MapR as a low-risk big data solution to build a digital oilfield. Efficiently integrated and analyzed the data to increase drilling performance and interpretation quality. Analyzed sensors and well log data in HDFS with HiveQL and prepare for prediction learning models. Constantly monitored the data and models to identify the scope of improvement in the processing and business. Manipulated and prepared the data for data visualization and report generation. Performed data analysis, statistical analysis, generated reports, listings and graphs. Co-leader of mathematics community 2015, Schlumberger Eureca. Data Scientist/Advanced Software Engineer TGS nopec R&D - Houston, TX December 2011 to December 2012 Using C/C++ programming language, develop and support modules of grid based model operations and acquisition geometry simulations. Support script based workflows, including I/O of large datasets, calling geophysical modules, and communications of parallel processes. Seismic attributes clustering using unsupervised neural network to help better utilizing big volume of geophysical data for geological feature recognition and oil production risk

control. Acquisition geometry simulation using Matlab/Unix Scripts. Seismic Data Processor CGGVeritas - Houston, TX July 2008 to December 2011 Seismic Data Processing & Interpretation; generated high quality subsurface images as a part of the team, with strong problem-solving ability, efficiency communication and corporation. Load acquisition data for processing with Oracle database management. Perform data analysis from different domains using both geophysical and statistical methods. Detect noisy/bad traces based on anomalous analysis. Predict and subtract noise models and surface related multiple models based on pattern reorganization, convolution methods and amplitude ratio identification. Clean/archive data sets on schedule. Processed seismic surveys of terabyte-datasets. Financial Analyst TPR Investments, L.P September 2006 to July 2008 Financial and Business analytics service using SAS and C. Generate prediction and regression model using statistical supervised learning methods. Collaborated with the analyst group to analyze the portfolio and evaluate the models. Estimate Value-at-Risk using Monte Carlo Simulation. Performed data analysis, statistical analysis, generated reports, listings and graphs. SAS financial time series autocorrelation analysis and regression using SAS/SGPLOT, PROC AUTOREG, with investigations into the model parameters AIC, AICC, MAPE, R-Square, etc. Trained the stock market prediction models and select the model with data mining methods. Selected Projects Unsupervised Neural Network in Geologic Feature Recognition Project -Clustering the seismic attributes into meaningful geological categories robustly. -Combination of principal component (PCA) and Self Organizing map (SOM). Artificial Intelligence Deep Learning Project -Lung Nodule Classification, ImageNet/GoogleNet, Object Detection by SSD and YOLO, Deep Speech. -Intel NEON/Tensorflow/Keras, performance evaluations. Job Salary Prediction Project -Process the text file with natural language processing tools-NLTK, word2vec, and efficiently predict the salary of job information posted online. - NLTK, Tensorflow NLP, openNLP/Java for feature preparation; regression model selected from Random Forest, Logistic Regression and Stochastic Gradient Descent(SGD) Deep Learning Application in Health Care - Lung Nodule Classification by 3D CNN, improved the mAP from 65% (base line) to 76%. - Spark BigDL/Tensorflow/Intel NEON/Keras. Cusomer Marketing Campaign and Risk Management

Project - Predict the charge-off event based on collected custom data for risk management in credit industry. - Compared XGboost, SVM, Multilayered Perceptron (MLP), and Random Forest. Education P h.D of Mathematics in Teaching Assistant Rice University - Houston, TX May 2006 MS in Applied Mathematics Huazhong University of Science& Technology - Wuhan, CN June 1998 Skills Python, Java, R, SAS, SQL, Hadoop, Hive, Spark, Tensorflow, bigDL, Machine Learning, Deep Learning (7 years), Python, Hadoop, Machine Learning, R, Spark Links https://github.com/yuewu000 Certifications/Licenses SAS Certified Base Programmer for SAS 9 Present SAS Certified Advanced Programmer for SAS 9 Present

Name: Aimee Rodriguez

Email: robin25@example.org

Phone: 5907596226