Yifan Wu

(206) 434 0079 | yw515@georgetown.edu

Summary: Experienced in differentiated insights extraction and visualization through Machine Learning, statistical method, and Big Data Analytics. Expert in creating industry-scalable product through model fine-tuning and AWS cloud operation. Rigorously trained to develop an acute sensitivity towards identifying key patterns in multidimensional data.

Experience:

EAGLEVIEW TECHNOLOGIES, Arlington, VA **May. 2018 – Aug 2018**
Data Scientist Intern, Machine Learning Team, Data Science Department

- Conducted exploratory analysis and thorough cleaning on property geospatial data for higher prediction accuracy
- Optimized the efficiency of AWS machine learning pipeline by 40% and generated prediction results for client
- Queried PostgresSQL database for previous ML model predictions and hand-selected true positive images from the query result for ML model training
- Fine-tuned Neural Network model Resnet 50 in AWS and increased model accuracy by 24%
- Designed and fully-automated the interactive visualization process for tracking ML team's project progression

GEORGETOWN UNIVERSITY, D.C. **Sep. 2017 – Dec. 2017**
Graduate Student, Project: Data Analysis for Insurance Rate and Hospital Expenditure

- Web-scrapped two data sets containing more than 15,000 records about hospital expenditure and insurance coverage rate from US Census Bureau
- Performed data normalization and identified two clusters by performing Hierarchical Clustering analysis with Python
- Proposed null hypothesis from the pattern uncovered by performing Association Rule text mining analysis
- Tested the null hypothesis using Support Vector Machine and Decision Tree to conclude that insurance rate is independent of geographic region, hospital expenditure, and county population

GEORGETOWN UNIVERSITY, D.C. **Sep. 2017 – Dec. 2017**
Graduate Student, Project: Sales Prediction for Weather-sensitive Products at Walmart Stores

- Collected datasets containing sale statistics for items sells in different Walmart stores and nearby weather station observations for the identification of weather sensitive product
- Eliminated anomalies and merged all datasets by shared attributes for later Machine Learning prediction
- Built Decision Tree Regression and Gradient Boosting Regression models to decide whether an item is a weather sensitive product and its future predicted sale under specific weather conditions
- Evaluated the performance of the models by calculating the area under the probability curve plotted with true positive rate against false positive rate (AUC of ROC curve)

UNIVERSITY OF WASHINGTON, Seattle, WA **Mar. 2016 – Jun. 2016**
Undergraduate Research Assistant, Michael G. Foster School of Business (Robert W. Palmatier)

- Documented data-breach records of 300 companies for categorizing companies' multi-faceted responses
- Uncovered a trend suggesting that companies' willingness to reveal data-breach incidents is inversely proportional to its size with Regression Analysis
- Conducted intense online research to compile comprehensive Excel database of the most prolific business and marketing textbook publishers to guide Dr. Palmatier's textbook-writing process
- Identified leading textbooks relating to marketing strategies, relationship marketing, customer loyalty, and relationship loyalty that enable Dr. Palmatier to consider the currently-available textbook content in preparing for drafting a proprietary business and marketing textbook
- Reviewed Dr. Palmatier's first draft of a textbook published in Feb. 2017 to identity and record all instances of domestic and international companies to enable Dr. Palmatier to advertise this textbook to foreign markets

Skills:

Languages: Python (TensorFlow, Pandas, Numpy, Boto3, MXNet, Scikit-Learn, Regular Expression, Natural Language Toolkit, Web-scraping, Text-mining, Hypothesis testing, Cross-Validation, ROC curve/AUC), R
Algorithms: Artificial Neural Network (Multi-layer Perceptron), Logistic Regression, Linear Regression, Lasso/Ridge Regularization, Decision Tree, KNN, Naïve Bayes, Ensemble Learning (Boosting, Bagging, Random Forests), Support Vector Machine, Clustering Analysis (K-means, Hierarchical, DBSCAN)
Cloud: S3, Hadoop, MapReduce, EMR, Hive, Pig, Spark, EC2, console page, DynamoDB
Visualization/Geospatial: Tableau, MatplotLib and Plotly, ggplot2, QGIS

Education:

GEORGETOWN UNIVERSITY, D.C., USA **Aug. 2017 - May. 2019**
Master of Science in Analytics, concentration: Data Science, GPA 3.7/4.0

- Course work: Neural Nets and Deep Learning, Data Analytics, Analytical Data Visualization, Probability Modeling/Statistical computing, Constrained and Unconstrained Convex Optimization, NLP for Data Analytics, Massive Data with AWS, Statistical Learning and Hypothesis Testing

UNIVERSITY OF WASHINGTON, Seattle, WA, USA **Aug. 2013 – Jun. 2017**
Bachelor of Art in Mathematics, GPA 3.5/4.0

- Course work: Calculus(I/II/III), Advanced Multivariable Calculus, Matrix Algebra with Applications, Linear Analysis, Real Analysis(I/II), Probability(I/II/III), Statistical Methods in Engineering and Science (R)