YING FANG

Concord, MA
fangyingalex@gmail.com
Mobile: (508)-873-6290

SKILLS

Programming Languages:	Python, R, SQL, Java, Shell Script, HTML5
Analytical Tools:	Numpy, Pandas, SciPy, Scikit-learn, TensorFlow, Keras, PySpark, AWS, BigQuery
Techniques:	Regression, Classification, Clustering, Text Mining, Neural Networks, A/B Testing
Software:	Tableau, MS SQL Server, Oracle Database, MySQL, SPSS, MS Access

WORK EXPERIENCE

Programming Analyst (Consultant)
SAI Global Compliance, Watertown MA **Nov. 2019 - Apr. 2020**

- Integrated data from multiple sources to support the Salesforce Customer **Data Migration** project.
- Built programs for accelerating **data manipulation** and **data wrangling** for 10k contract data using **Python**.
- Supported Finance Team on the **Deferred Revenue Analysis** to examine irregularities using Python, Excel.
- Implemented dashboards in **Tableau** for customer segmentation in **K-means** and sales summary reporting.

Data Science Co-op
John Hancock Financial Services, Boston MA **Jan. - Sept. 2019**

- **Client Retention Project** - Developed predictive models (**Logistic Regression**, **Decision Trees**) in **Python** to identify clients with a high risk of termination. Delivered the predictions and analysis in **Tableau** dashboards. The tagged clients are 2.94 times more likely to discontinue in the next 6 months.
- Designed **utilization metrics**, generated analytical reports to support business decision making.
- Improved the analysis reporting process by implementing **data automation** using **SQL** and **Python**.
- Implemented visualizations in **Tableau** using **SQL Server** as data sources for ad-hoc requests.
- Built customized **data pipelines** using Python from Database and **RESTful API** for higher data availability.
- Constructed **Bi-LSTM-CNNs** + **Regex** for Personally Identifiable Information detection with 72% accuracy.

Data Science Summer Intern
China Organizational Name Administration Center, Beijing, China **May - Aug. 2018**

- Constructed an analytical pipeline for Online Public Opinion Hotspot Detection and Analysis.
  Cleaned raw data from web crawlers, segmented Chinese words and applied **Word2Vec** for word embedding.
  Built **Sentiment Classifiers** with SVM, CNN and LSTM. Achieved an accuracy rate of 94% with LSTM.

EDUCATION

Worcester Polytechnic Institute (WPI), Worcester, MA **May 2019**
Master of Science in **Data Science** | GPA: 3.63/4.00

Sichuan University, Chengdu, China **June 2016**
Bachelor of Management in **Industrial Engineering** | GPA: 3.39/4.00

PROJECTS

Google Analytics Customer Revenue Prediction with Python, Kaggle **Oct. - Dec. 2018**

- Implemented a machine learning system to identify potential customers and predict their spending.
- Built a pre-classified regression model to predict customer's spending with Scikit-learn based on an extremely skewed dataset (< 1.3% target feature are non-zero) and reduced the loss (RMSE) from 1.84 to 1.67.
- Constructed a **weighted RNN** model to predict revenue generated by customers who visited more than one time. Reduced the loss (RMSE) from 2.2 to 1.9.

Bank Telemarketing Success Prediction with R, WPI **Oct. - Nov. 2017**

- Preprocessed data with over-sampling, feature standardization and dimension reduction using **PCA**.
- Performed classification models including **Logistic Regression**, **Naïve Bayes**, **Decision Trees** and **SVM** on the training set with **confusion matrices**. Evaluated the best model of LR with an accuracy of 89.4%.