

# Ruxin Li

(757) 837-1542 | [rli11@email.wm.edu](mailto:rli11@email.wm.edu) | [linkedin.com/in/ruxin-li](https://www.linkedin.com/in/ruxin-li) | [ruxinli.github.io](https://ruxinli.github.io)

---

## SUMMARY

Strong data analytical professional with Supply Chain background and practical data-related project experience. Creative, passionate, organized team leader motivated to solve company problems through analyzing big data; able to communicate complex concepts to diverse audiences.

## SKILLS

Programming Tool: Python (Sklearn, Pandas, Numpy), R, SQL, Excel VBA

Machine Learning: Bayes' theorem, GLM, Tree-based Methods (XGBoost, Decision Tree, Random Forest), SVM, Principle Component Analysis, Clustering

Big data tool: Pyspark, MapReduce, AWS S3, SageMaker

Visualization: Tableau, Python (Seaborn, Matplotlib)

## EDUCATION

<b>William and Mary, Raymond A. Mason School of Business</b>	May 2020
Master of Business Analytics	GPA:3.82

<b>Old Dominion University</b>	Dec 2018
Bachelor of Science, Supply Chain Management	GPA:3.71

Bachelor of Science, Business Analytics

## EXPERIENCE

### Ferguson

*Modeling Analyst*

Feb 2020 - May 2020[[link](#)]

- Designed automated data pipelines for over 40000 rows of integrated data (external and internal) using Alteryx and MySQL manipulated and cleaned data with Pandas in Python
- Recommended new showroom locations for Ferguson by developing predictive analytics models (GBM, XGBoost, and SVM, etc.) in R, evaluated the models by 10-fold validation, selected the best model based on MAE
- Produced interactive dashboard in Tableau to provide location insights for Enterprise Information Management Team

### Deppon Logistics

*Logistic data statistician*

Jan 2019- June 2019

- Optimized logistics transport capacity of a cosmetic company customer by building binary constraint model in Excel Open Solver
- Created a visual dashboard of monthly transportation volume, periods delay of logistics nodes, and presented to managers to decide major routes of customers, saved transportation space 16.7%, transportation cost \$69,640

### SYNUTRA France International

*International trade data statistician*

June 2019- August 2019

- Classified diverse dairy products into groups, reallocated cargo space based on Best Fit Algorithm in Python
- Provided forecasts of weekly container stuffing charge by building ARIMA Time Series model in R
- Wrote reports about stuffing charges prediction in 95% confidential intervals, reduced goods backlogs by 20%

## PROJECTS

---

**Conversion Rate prediction using AWS SageMaker** Mar 2020- April 2020[\[link\]](#)

- Managed, accessed over 300,000 records in S3 bucket, built machine learning models in notebook instance using AWS SageMaker
- Trained and tested the models by customizing a 6-core, Linux-based machine in AWS EC2
- Predict user conversion rate for an e-commerce company based on Gradient Boosting classifier framework, produced an overall accuracy of 98.6%

**News Headline Prediction with RNN** Mar 2020- April 2020[\[link\]](#)

- Transformed, trimmed the text data, mapped unique words to integers in Python deep learning API Keras and Sklearn
- Built long shot-term memory (LSTM) Neural Network model with a loss function of categorical cross entropy, set model checkpoint to monitor the loss
- Tuned the Neural Network model by changing the number of units, layers and prediction sequence length.
- Generated reasonable text prediction for news headlines.

**Big Data Analytics: Airbnb Top Cost-Effective Houses** Dec 2019- Feb 2020[\[link\]](#)

- Processed, transformed Airbnb housing data into RDD using Big Data platform PySpark
- Wrote SQL queries in Spark to perform analytical tasks, provided sorted lists of most cost-effective houses on Airbnb

**Kobe Shot Prediction** Jan 2020- Feb 2020[\[link\]](#)

- Developed and compared machine learning models (Random Forest, Logistic Regression, XGBoost, SVC etc.) in R to predict whether the basket went in given a shot of Kobe
- Utilized PCA to select features, implemented Stacked Ensemble Method in machine learning to combine different models
- Evaluated the models using 10-fold validation and model correlation, selected and tuned the best model

## ADDITIONAL INFORMATION

---

### *Leadership*

- Minister of Organizing Department, CSSA
- Captain, campus Women's Soccer Team

**Volunteer:** Union Mission Ministries, organized CSSA members to sort and carry out donation materials

**Languages:** Fluent in English and Chinese

**Certifications:** Beta Gamma Sigma, Starbucks Barista

**Interests:** Soccer, Jazz and Breaking, Extreme Sports

**Personal Website:** [ruxinli.github.io](https://ruxinli.github.io)