YIMEI FAN

1105 W NC Highway 54 Bypass, Apt. F11, Chapel Hill, North Carolina, the United States 27516
+1 (404) 421-3025   yimei@live.unc.edu

SKILLS

Find my work here: https://github.com/Yimei120

Personal Website: https://yimei120.github.io/YimeiFan.github.io/

Languages: Python (Pandas, NumPy, Matplotlib), Java, R (dplyr, ggplot2)

Databases: MySQL, SQL Server

Software: Tableau, MS Excel

Statistics: A/B testing, Hypothesis Testing, Linear Regression, Logistic Regression, Naïve Bayes, Lasso Regression, Ridge Regression

WORK EXPERIENCES

Amazon Co., Ltd.   September 2018 – June 2019

Business Analyst Intern

- Designed and developed **ETL** structures using **SQL**, including 6 data feeding jobs, 9 metrics, and 1 consolidation job, which could be utilized by India, Japan and China teams, for transformation of data sources into data warehouses.
- Constructed a metric as well as ETL jobs to monitor abnormal return orders automatically, and improved the customer return process, which used to be completed manually and inefficiently.
- Built **Tableau** dashboards regarding to "Expected Shipping Date (ExSD)"to monitor performances of vendors and improved vendors' Estimated Delivery Accuracy (EDA) by 300bps within one week.
- Supported daily data requests, maintained 11 dashboards at the same time, and addressed more than 200 stuck order issues through troubleshooting and debugging during a two-week promotion peak in both China and Japan.

Schneider Electric Co., Ltd.   January 2018 - May 2018

Data Cleaning Intern

- Provided trainings for 3 new team members and helped them get familiar with daily operations within one week, which usually took two weeks before.
- Selected and categorized data in transaction records of more than 200 clients per business day.
- Analyzed database designs and system integration process utilizing **MS Excel** pivots, and made suggestions for further enhancement.

EDUCATION

University of North Carolina at Chapel Hill   August 2019 – May 2021

Master of Science in Information Science

Coursework: Data Structure; Machine Learning; Information Retrieval; Database Concepts& Applications

Beijing International Studies University (BISU)   September 2015 – June 2019

Bachelor of Economics, Major in Trade Economics   GPA: 3.73/4.00; Top 10%

PROJECTS EXPRERIENCES

Explorational West Nile Virus Classification   August 2019- December 2019

Team Leader

- Employed data cleaning, data merging and data visualization to achieve explorational data analysis, preparing for classification model fitting later on.
- Fitted models including Naïve Bayes Classifier, K-nearest Neighbor, logistic regression, classification tree, gradient boosting, and random forest models, using both confusion matrix and ROC curve to evaluate models.
- Logistic regression model is the best for West Nile Virus classification, which has testing accuracy as 93.96% and better ROC curve.

Carolina Data Challenge – Crime Prediction by Machine Learning   October 5th, 2019- October 6th, 2019

Sponsored by EY, SAS, and CapTech

- Accomplished a data analysis project regarding to crime category prediction within 24 hours by employing real-life datasets of crimes in Chapel Hill, including *Police Incidents Reports Written* and *Police Arrests Made.*
- Utilized **Python Pandas** and **NumPy** to index and clean databases, and then employed **Python Matplotlib** and **Tableau** to construct histograms, correlation matrix and distribution maps to analyze the trend and observe the distribution of raw data.
- Constructed a **K-Nearest-Neighbor** model to predict potential crime categories with chosen variables, such as gender, race, age, or ethnicity, and the highest accuracy of the model is 40.52% when k equals to 40.
- The crime category the model could predict most accurately was DWI (Driving While Intoxicated). With this model, police could be able to prevent this kind of crime given particular human characteristics.

Using Machine Learning to Improve Destination Management Organizations' Photo Selection   March 2018 – September 2018

Advised by Professor Ning Deng (Beijing International Studies University)

- Extracted useful contents from raw data, which included photos and texts, and analyzed **textual data**, specifically, calculating frequency of adjective and nouns to extract true attitudes of tourists using **Textblob**, a Python library that offers a simple API to access its methods and performs basic **Natural Language Processing (NLP)** tasks.
- Generated textual descriptions of photos, which contained pairs of nouns and adjectives using **DeepSentiBank**, a visual sentiment concept classification method based on deep convolutional neural networks.
- Employed a **classifier algorithm** to predict the correlation between user-generated contents and machine-generated contents to evaluate the accuracy of machine-generated contents in terms of whether it can present people's opinions.