Di Wu

206-604-6002 | di.wu@ur.rochester.edu | www.linkedin.com/in/diwuethan

EDUCATION

SIMON BUSINESS SCHOOL - University of Rochester – Rochester, NY July 2019 - Present

Master of Science in Business Analytics (STEM-certified) - GPA: 3.6/4 *Available for full time work in May. 2020*

Scholarship: 15% Merit Scholarship recipient

University of Washington - Seattle, Washington Sep 2015 - June 2019

Bachelor of Economics, Concentration in Econometrics - GPA: 3.5/4

INTERNSHIP EXPERIENCE

Parisa Wang - luxurious handbag designer 01/2020 – Present

Business Intelligence Analyst New York, New York

- Computed market research by web crawling attributes of 1200+ products with python BeautifulSoup and Selenium Web Driver; compared Parisa Wang's price and product attributes across different competitors to support future pricing strategy.
- Created over 100 SQL queries to extract customer click activities and traffic channels from company database; designed data extraction framework with 120+ lines of SQL queries which achieved reduction of data-processing time by 60%.
- Computed Funnel Analysis to visualize customer web journey by device. Formulated actionable recommendations on webpage optimization based on conversion rate performance, increased Key performance indicator by 25%.
- Visualized analysis results by creating 10 dashboards in Tableau for presentation to management and stockholders of Parisa Wang; acquired 150 waitlist subscribers and converted 50 handbag pre-sales within three weeks during pre-launch stage.

Catholic Family Center - family services provider 02/2020 – Present

Data Analyst Intern Rochester, New York

- Cleaned admission & discharge datasets which contains over 26,000 patient records with 150+ variables using R; Applied feature selection by checking correlation score and Principal component analysis, achieved 30% redundant dimension reduction.
- Conducted descriptive analysis and patient segmentation based on clinical and demographic data; Visualized clustering outcome of patient trajectories and analyzed effect of program participation to identify possible program improvements.
- Developed machine learning models to help CFC better allocate patients by predicting program performance and no-show rate for counselling session; Enhanced potential program performance by 20% and recommended results to stakeholders.

PROJECT EXPERIENCE

Movie Recommendation Engine Development 02/2020 – 03/2020

- Built data ETL pipeline to analyze movie rating dataset and constructed online analytical processing with Spark SQL.
- Implemented the Alternative Least Square model to provide personalized movie recommendations and developed user-based approaches to handle system cold-start problems; Conducted model hyper-parameters tuning with Spark ML cross-evaluation toolbox and monitor data performance via Spark UI on AWS.

Pricing Strategy for Soft Drink Company 12/2019 – 01/2020

- Built Multi-Logit model using R based on over 8000 purchasing records; estimated aggregated demand for existing products.
- Identified six potential segments of targeted customers based on 13 demographic features of targeted customers using K-means; computed targeting price based on segments specified demand function which increased profit by 8%.

Negative Tweet Detection for Airlines 10/2019 – 11/2019

- Conducted sentimental analysis on 4000 tweets to predict negative comments for an Airline company; preprocessed data by tokenization and converted text to features by Term Frequency – Inverse Document Frequency (TFIDF).
- Built machine learning models and evaluated the model's performance via 10-fold cross validation; selected Gradient boost based on best AUC of 0.875 and predicted the possibility of negative tweet comments for airlines (~22%).

SKILLS

- **Programming**: Python (Sklearn, Pandas, NumPy, Seaborn, Beautiful soup, Matplotlib), R (ggplot2, tidyverse, sentimentr, plotly, stringr), SQL (T-SQL, MySQL, Spark), Microsoft Excel, Tableau, Jupyter Notebook.
- **Machine Learning:** Principal Components Analysis (PCA), K-means, Logistic regression, Classical & Penalized Regression Methods (Lasso, Ridge), K Nearest Neighbors, Random forest, Native Bayes, XgBoosts, Gradient Boost, ARIMA.
- **Statistics analysis:** A/B testing, Chi-Square testing, Hypothesis testing, time-series analysis.
- **Certification:** Advanced Google Analytics individual Qualification, Machine Learning with Python,